

DCR-CORE Processing Results

Level	Container	Tag	opt	Description	Type	Source
10	document	document		container for document	root	TET
11	document	fileName		file name	str	TET
11	document	id	yes	document identification	int	TET
11	document	noFonts		number fonts	int	TET
11	document	noLines		number lines in document	int	TET
11	document	noLinesFooter	yes	number footer lines	int	TET
11	document	noLinesHeader	yes	number header lines	int	TET
11	document	noLinesToc	yes	number toc lines	int	TET
11	document	noListsBullet	yes	number bulleted lists	int	TET
11	document	noListsNumber	yes	number numbered lists	int	TET
11	document	noPages		number pages in document	int	TET
11	document	noParas		number paragraphs in document	int	TET
11	document	noSentences	yes	number sentences in document	int	spaCy
11	document	noTables	yes	number tables in document	int	TET
11	document	noTitles	yes	number titles in document	int	TET
11	document	noWords		number words in document	int	TET
19	document	containerFonts		container for fonts	list	TET
19	document	containerPages		container for pages	list	TET
19	document	containerTitles	yes	container for titles	list	TET
20	font	font		container for font	root	TET
21	font	embedded		embedded	bool	TET
21	font	fontNo		font number in document	int	TET
21	font	fullName		font full name	str	TET
21	font	id		font identification	str	TET
21	font	italicAngle		font italic angle	float	TET
21	font	name		font name	str	TET
21	font	type		font type	str	TET
21	font	weight		font weight	float	TET
30	page	page		container for page	root	TET
31	page	lineNoFirst		number of first line in page	int	TET
31	page	lineNoLast		number of last line in page	int	TET
31	page	pageNo		page number in document	int	TET
31	page	paraNoFirst		number of first paragraph in page	int	TET
31	page	paraNoLast		number of last paragraph in page	int	TET
31	page	sentenceNoFirst		number of first sentence in page	int	TET
31	page	sentenceNoLast		number of last sentence in page	int	TET
31	page	wordNoFirst		number of first word in page	int	TET
31	page	wordNoLast		number of last word in page	int	TET
39	page	lines		container for lines	list	TET
39	page	paras		container for paragraphs	list	TET
39	page	sentences		container for sentences	list	spaCy
40	line	line		container for lines	root	TET
41	line	lineNo		line number in document	int	TET

DCR-CORE Processing Results

Level	Container	Tag	opt	Description	Type	Source
41	line	lineNoPage		line number in page	int	TET
41	line	lineNoPara		line number in paragraph	int	TET
41	line	llx		x coordinate of the lower left corner	float	TET
41	line	text		line text	str	TET
41	line	type		line type	str	TL
41	line	urx		x coordinate of the upper right corner	float	TET
41	line	wordNoFirst		number word of first word in line	int	TET
41	line	wordNoLast		number word of last word in line	int	TET
50	para	para		container for paragraph	root	TET
51	para	lineNoFirst		number of first line in paragraph	int	TET
51	para	lineNoLast		number of last line in paragraph	int	TET
51	para	paraNo		paragraph number in document	int	TET
51	para	paraNoPage		paragraph number in page	int	TET
51	para	sentenceNoFirst		number of first sentence in paragraph	int	TET
51	para	sentenceNoLast		number of last sentence in paragraph	int	TET
51	para	text		paragraph text	str	TET
51	para	wordNoFirst		number of first word in paragraph	int	TET
51	para	wordNoLast		number of last word in paragraph	int	TET
59	para	words		container for words	list	TET
60	sentence	sentence		container for sentences	root	TET
61	sentence	sentenceNo		sentence number in document	int	TET
61	sentence	sentenceNoPage		sentence number in page	int	TET
61	sentence	sentenceNoPara		sentence number in paragraph	int	TET
61	sentence	text		sentence text	str	TET
61	sentence	wordNoFirst		number word of first word in sentence	int	TET
61	sentence	wordNoLast		number word of last word in sentence	int	TET
70	word	word		container for word	root	TET
71	word	font		font identification	str	TET
71	word	size		font size	float	TET
71	word	text		word text	str	TET
71	word	tknCluster			int	spaCy
71	word	tknDep_			str	spaCy
71	word	tknDoc			str	spaCy
71	word	tknEntIob_			str	spaCy

DCR-CORE Processing Results

Level	Container	Tag	opt	Description	Type	Source
71	word	tknHead			int	spaCy
71	word	tknI			int	spaCy
71	word	tknIdx			int	spaCy
71	word	tknIsAlpha			bool	spaCy
71	word	tknIsAscii			bool	spaCy
71	word	tknIsOov			bool	spaCy
71	word	tknIsSentStart			bool	spaCy
71	word	tknIsTitle			bool	spaCy
71	word	tknLang_			str	spaCy
71	word	tknLeftEdge			int	spaCy
71	word	tknLemma_			str	spaCy
71	word	tknLex			str	spaCy
71	word	tknLexId			int	spaCy
71	word	tknLower_			str	spaCy
71	word	tknMorph			str	spaCy
71	word	tknNorm_			str	spaCy
71	word	tknOrth_			str	spaCy
71	word	tknPos_			str	spaCy
71	word	tknPrefix_			str	spaCy
71	word	tknProb			float	spaCy
71	word	tknRank			int	spaCy
71	word	tknRightEdge			int	spaCy
71	word	tknSent			str	spaCy
71	word	tknSentiment			float	spaCy
71	word	tknShape_			str	spaCy
71	word	tknSuffix_			str	spaCy
71	word	tknTag_			str	spaCy
71	word	tknText			str	spaCy
71	word	tknTextWithWs			str	spaCy
71	word	tknVocab			str	spaCy
71	word	tknWhitespace_			str	spaCy
71	word	wordNo		word number in document	int	TET
71	word	wordNoLine		word number in line	int	TET
71	word	wordNoPage		word number in page	int	TET
71	word	wordNoPara		word number in paragraph	int	TET
100	heading	toc	yes	container for toc	root	LT
101	heading	entries		container for heading entries	list	LT
110	entry	entry		container for heading entry	root	LT
111	entry	ctxtLine1		1. context line	str	LT
111	entry	ctxtLine2		2. context line	str	LT
111	entry	ctxtLine3		3. context line	str	LT
111	entry	level		heading level	int	LT
111	entry	lineNoPage		line number in page	int	LT
111	entry	pageNo		page number in document	int	LT
111	entry	regexp	yes	regular expression	str	LT
111	entry	text		heading text	str	LT
200	lists bullet	listsBullet	yes	container for bulleted lists	root	LT
201	lists bullet	lists		container for bulleted lists	list	LT
210	list	list		container for list	root	LT

DCR-CORE Processing Results

Level	Container	Tag	opt	Description	Type	Source
211	list	entries		container for list entries	list	LT
211	list	format		number format	str	LT
211	list	listNo		list number in document	int	LT
211	list	noEntries		number entries in list	int	LT
211	list	pageNoFirst		page number first entry	int	LT
211	list	pageNoLast		page number last entry	int	LT
220	entry	entry		container for bulleted list entry	root	LT
221	entry	entryNo		entry number in list	int	LT
221	entry	lineNoPageFirst		first line in page	int	LT
221	entry	lineNoPageLast		last line in page	int	LT
221	entry	pageNo		page number in document	int	LT
221	entry	paraNo		paragraph number in page	int	LT
221	entry	text		list entry text	str	LT
300	lists number	listsNumber	yes	container for numbered lists	root	LT
301	lists number	lists		container for numbered lists	list	LT
310	list	list		container for list	root	LT
311	list	format		bullet format	str	LT
311	list	listNo		list number in document	int	LT
311	list	noEntries		number entries in list	int	LT
311	list	pageNoFirst		page number first entry	int	LT
311	list	pageNoLast		page number last entry	int	LT
311	list	regexp	yes	regular expression	str	LT
319	list	entries		container for list entries	list	LT
320	entry	entry		container for numbered list entry	root	LT
321	entry	entryNo		entry number in list	int	LT
321	entry	lineNoPageFirst		first line in page	int	LT
321	entry	lineNoPageLast		last line in page	int	LT
321	entry	pageNo		page number in document	int	LT
321	entry	paraNo		paragraph number in page	int	LT
321	entry	text		list entry text	str	LT
400	tables	tables	yes	container for document tables	root	LT
401	tables	tables		container for document tables	list	LT
410	table	table		container for document table	root	LT
411	table	llxFirstRow		first row: x coordinate of the lower left corner	float	LT
411	table	noColumns		number columns in table	int	LT
411	table	noRows		number rows in table	int	LT
411	table	pageNoFirst		page number first row	int	LT
411	table	pageNoLast		page number last row	int	LT
411	table	tableNo		table number in document	int	LT
411	table	urxFFirstRow		first row: x coordinate of the upper right corner	float	LT
419	table	rows		container for table rows	list	LT
420	row	row		container for table rows	root	LT
421	row	llxFirstColumn		first column: x coordinate of the lower left corner	float	LT

DCR-CORE Processing Results

Level	Container	Tag	opt	Description	Type	Source
421	row	rowNo		number row in table	int	LT
421	row	urxFirstColumn		first column: x coordinate of the upper right corner	float	LT
429	row	columns		container for table columns	list	LT
430	column	column		container for table column	root	LT
431	column	columnNo		number column in row	int	LT
431	column	lineNoPage		line number in page	int	LT
431	column	llx		x coordinate of the lower left corner	float	LT
431	column	paraNo		paragraph number in page	int	LT
431	column	text		column text	str	LT
431	column	urx		x coordinate of the upper right corner	float	LT