msaSDK Module

`.models.sdu`

Module for the Semantic Document Understanding - Content

Classes

SDUAttachment

Bases: `SQLModel`

Attributes

binary `class-attribute`

```
binary: bool = False
```

charset `class-attribute`

```
charset: str = ''
```

content_type `class-attribute`

```
content_type: str = ''
```

disposition `class-attribute`

```
disposition: str = ''
```

encoding `class-attribute`

```
encoding: str = ''
```

id `class-attribute`

```
id: str = ''
```

metadata class-attribute

```
metadata: Dict = {}
```

name class-attribute

```
name: str = ''
```

path class-attribute

```
path: str = ''
```

payload class-attribute

```
payload: str = ''
```

status class-attribute

```
status: str = ''
```

text class-attribute

```
text: SDUText = SDUText()
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUBBox

Bases: `SQLModel`

Attributes

x0 class-attribute

```
x0: float = -1
```

x1 class-attribute

```
x1: float = -1
```

y0 class-attribute

```
y0: float = -1
```

y1 class-attribute

```
y1: float = -1
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUContent

Bases: `SQLModel`

Attributes

attachments class-attribute

```
attachments: List[SDUAttachment] = []
```

layouts class-attribute

```
layouts: List[SDULayout] = []
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUData

Bases: SQLModel

Attributes

converter class-attribute

```
converter: List[str] = []
```

email class-attribute

```
email: SDUEmail = SDUEmail()
```

images class-attribute

```
images: List[SDUPageImage] = []
```

npages class-attribute

```
npages: int = 0
```

pages class-attribute

```
pages: List[SDUPage] = []
```

stats class-attribute

```
stats: SDUStatistic = SDUStatistic()
```

text class-attribute

```
text: SDUText = SDUText()
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

Functions

addPagePreProcessing

```
addPagePreProcessing(pagepre: SDUPage)
```

escaped

```
escaped()
```

sanitized async

```
sanitized()
```

SDUDimensions

Bases: `SQLModel`

Attributes

factor_x class-attribute

```
factor_x: float = 0.0
```

factor_y class-attribute

```
factor_y: float = 0.0
```

height class-attribute

```
height: float = 0.0
```

id class-attribute

```
id: int = -1
```

rotation class-attribute

```
rotation: int = 0
```

width class-attribute

```
width: float = 0.0
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUElement

Bases: `SQLModel`

Attributes

end class-attribute

```
end: int = -1
```

id class-attribute

```
id: int
```

start class-attribute

```
start: int = -1
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUEmail

Bases: SQLModel

Parsed EMail Pydantic Model.

Attributes

msg_bcc class-attribute

```
msg_bcc: str = ''
```

msg_body class-attribute

```
msg_body: str = ''
```

msg_cc class-attribute

```
msg_cc: str = ''
```

msg_from class-attribute

```
msg_from: str = ''
```

msg_headers class-attribute

```
msg_headers: Dict = {}
```

msg_id class-attribute

```
msg_id: str = ''
```

msg_received class-attribute

```
msg_received: List = []
```

msg_reply_to class-attribute

```
msg_reply_to: str = ''
```

msg_sender_ip class-attribute

```
msg_sender_ip: str = ''
```

msg_sent_date class-attribute

```
msg_sent_date: str = ''
```

msg_subject class-attribute

```
msg_subject: str = ''
```

msg_timezone class-attribute

```
msg_timezone: str = ''
```

msg_to class-attribute

```
msg_to: str = ''
```

msg_to_domains class-attribute

```
msg_to_domains: str = ''
```

seg_body class-attribute

```
seg_body: str = ''
```

seg_sign class-attribute

```
seg_sign: str = ''
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUFonts

Bases: `SQLModel`

Attributes

avg_fontsize class-attribute

```
avg_fontsize: int = 14
```

fonts class-attribute

```
fonts: List = []
```

fontsizes class-attribute

```
fontsizes: Dict = {}
```

id class-attribute

```
id: int = -1
```

small_fontsize class-attribute

```
small_fontsize: int = 10000
```

Classes

Config

Attributes

> orm_mode class-attribute
>
> ```
> orm_mode = False
> ```

SDULanguage

> Bases: `SQLModel`
>
> Detected Language Pydantic Model.
>
> **Attributes**
>
> bytes class-attribute
>
> ```
> bytes: int = -1
> ```
>
> code class-attribute
>
> ```
> code: str = 'unknown'
> ```
>
> confidence class-attribute
>
> ```
> confidence: float = -1
> ```
>
> details class-attribute
>
> ```
> details: Optional[Tuple] = tuple()
> ```
>
> lang class-attribute
>
> ```
> lang: str = 'unknown'
> ```
>
> proportion class-attribute
>
> ```
> proportion: int = -1
> ```
>
> reliable class-attribute
>
> ```
> reliable: bool = False
> ```
>
> winner class-attribute

```
winner: Optional[str] = None
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDULayout

Bases: SQLModel

Attributes

bjson class-attribute

```
bjson: Dict = {}
```

blocks class-attribute

```
blocks: List[tuple] = []
```

body class-attribute

```
body: SDUBBox = SDUBBox()
```

columns class-attribute

```
columns: List[SDUBBox] = []
```

dimensions class-attribute

```
dimensions: SDUDimensions = SDUDimensions()
```

drawings class-attribute

```
drawings: List = []
```

fonts class-attribute

```
fonts: SDUFonts = SDUFonts()
```

footer class-attribute

```
footer: SDUBBox = SDUBBox()
```

header class-attribute

```
header: SDUBBox = SDUBBox()
```

id class-attribute

```
id: int = -1
```

images class-attribute

```
images: List = []
```

layouts class-attribute

```
layouts: List = []
```

margin_left class-attribute

```
margin_left: SDUBBox = SDUBBox()
```

margin_right class-attribute

```
margin_right: SDUBBox = SDUBBox()
```

rows class-attribute

```
rows: List[SDUBBox] = []
```

texttrace class-attribute

```
texttrace: List = []
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDULearnset

Bases: `SQLModel`

Attributes

emb class-attribute

```
emb: Dict = {}
```

nlp class-attribute

```
nlp: Dict = {}
```

nlu class-attribute

```
nlu: Dict = {}
```

text class-attribute

```
text: Dict = {}
```

vec_sent class-attribute

```
vec_sent: Dict = {}
```

vec_words class-attribute

```
vec_words: Dict = {}
```

version class-attribute

```
version: str = ''
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

Functions

reset

```
reset()
```

set_version

```
set_version(version: str)
```

SDUPDFElement

Bases: SQLModel

Attributes

bold class-attribute

```
bold: bool = False
```

color class-attribute

```
color: int = 0
```

flags class-attribute

```
flags: int = 0
```

font class-attribute

```
font: str = ''
```

fontsize class-attribute

```
fontsize: float = 0.0
```

italic class-attribute

```
italic: bool = False
```

line_id class-attribute

```
line_id: int = -1
```

span_id class-attribute

```
span_id: int = -1
```

SDUPage

Bases: `SQLModel`

Attributes

has_en class-attribute

```
has_en: bool = False
```

input class-attribute

```
input: str = ''
```

npar class-attribute

```
npar: int = 0
```

page class-attribute

```
page: int = -1
```

text class-attribute

```
text: SDUText = SDUText()
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

Functions

getAllSentencesTextList

```
getAllSentencesTextList()
```

getAllSentencesTextListLF

```
getAllSentencesTextListLF()
```

getAllSentencesTextListNoTableAndLists

```
getAllSentencesTextListNoTableAndLists()
```

getAllSentencesTextList_en

```
getAllSentencesTextList_en()
```

getTextDefault

```
getTextDefault()
```

getTextForDisplay

```
getTextForDisplay()
```

getTextForNLP

```
getTextForNLP()
```

getTextLF

```
getTextLF(space_before_lf: bool = False)
```

getTextLF_Paragraph

```
getTextLF_Paragraph(space_before_lf: bool = False)
```

getTextNoLF

```
getTextNoLF()
```

getTextNoLF_EN

```
getTextNoLF_EN()
```

getTextNoLF_Paragraph

```
getTextNoLF_Paragraph()
```

hasText

```
hasText()
```

setInput

```
setInput(inputText: str)
```

SDUPageImage

Bases: `SQLModel`

Page Image Pydantic Model.

Storing the information about the Image representation of a Page.

Attributes

dpi `class-attribute`

```
dpi: float = 0.0
```

filepath_name class-attribute

```
filepath_name: str = ''
```

format class-attribute

```
format: str = ''
```

height class-attribute

```
height: float = 0.0
```

id class-attribute

```
id: int = -1
```

layout class-attribute

```
layout: List = []
```

mode class-attribute

```
mode: str = ''
```

width class-attribute

```
width: float = 0.0
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUParagraph

Bases: `SQLModel`

Attributes

clean class-attribute

```
clean: str = ''
```

elements class-attribute

```
elements: List[SDUPDFElement] = []
```

id class-attribute

```
id: int = -1
```

lang class-attribute

```
lang: SDULanguage = SDULanguage()
```

nsen class-attribute

```
nsen: int = 0
```

section class-attribute

```
section: str = 'body'
```

semantic_type class-attribute

```
semantic_type: str = 'text'
```

sentences class-attribute

```
sentences: List[SDUSentence] = []
```

sentences_en class-attribute

```
sentences_en: List[SDUSentence] = []
```

size_type class-attribute

```
size_type: str = 'body'
```

sort class-attribute

```
sort: int = -1
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

Functions

getText

```
getText() -> str
```

getTextLF

```
getTextLF() -> str
```

getTextNoLF

```
getTextNoLF() -> str
```

hasText

```
hasText() -> bool
```

SDUSentence

Bases: `SQLModel`

Attributes

id class-attribute

```
id: int = -1
```

text class-attribute

```
text: str = ''
```

tokens class-attribute

```
tokens: List[str] = []
```

upos class-attribute

```
upos: List[str] = []
```

xpos class-attribute

```
xpos: List[str] = []
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUStatistic

Bases: `SQLModel`

Text Statistics Pydantic Model.

Attributes

avg_character_per_word class-attribute

```
avg_character_per_word: float = 0
```

avg_letter_per_word class-attribute

```
avg_letter_per_word: float = 0
```

avg_sentence_length class-attribute

```
avg_sentence_length: float = 0
```

avg_sentence_per_word class-attribute

```
avg_sentence_per_word: float = 0
```

avg_syllables_per_word class-attribute

```
avg_syllables_per_word: float = 0
```

coleman class-attribute

```
coleman: float = 0
```

crawford class-attribute

```
crawford: float = 0
```

difficult_words class-attribute

```
difficult_words: int = 0
```

fog class-attribute

```
fog: float = 0
```

grade class-attribute

```
grade: float = 0
```

gulpease_index class-attribute

```
gulpease_index: float = 0
```

lexicon_count class-attribute

```
lexicon_count: int = 0
```

long_word_count class-attribute

```
long_word_count: int = 0
```

osman class-attribute

```
osman: float = 0
```

paragraph_count class-attribute

```
paragraph_count: int = 0
```

reading_ease class-attribute

```
reading_ease: str = ''
```

reading_ease_score class-attribute

```
reading_ease_score: float = 0
```

reading_index class-attribute

```
reading_index: float = 0
```

reading_score class-attribute

```
reading_score: float = 0
```

reading_time_s class-attribute

```
reading_time_s: float = 0
```

sentence_count class-attribute

```
sentence_count: int = 0
```

smog class-attribute

```
smog: float = 0
```

standard class-attribute

```
standard: str = ''
```

write_formula class-attribute

```
write_formula: float = 0
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUText

Bases: `SQLModel`

Attributes

clean class-attribute

```
clean: str = ''
```

html_content class-attribute

```
html_content: str = ''
```

lang class-attribute

```
lang: SDULanguage = SDULanguage()
```

paragraphs class-attribute

```
paragraphs: List[SDUParagraph] = []
```

raw class-attribute

```
raw: str = ''
```

structured_content class-attribute

```
structured_content: Dict = {}
```

Classes

Config

Attributes

orm_mode class-attribute

```
orm_mode = False
```

SDUVersion

Bases: `SQLModel`

Attributes

creation_date class-attribute

```
creation_date: str = ''
```

version class-attribute

```
version: str = ''
```

Classes

Config

Attributes

> orm_mode `class-attribute`
>
> ```
> orm_mode = False
> ```

Functions

getCRLF

```
getCRLF() -> str
```

> get's the OS Environment Variable for `CR_LF`. Default: `\n`

getCRParagraph

```
getCRParagraph() -> str
```

> get's the OS Environment Variable for `CR_PARAGRAPH`. Default: `\n\n`

getSentenceSeperator

```
getSentenceSeperator() -> str
```

> get's the OS Environment Variable for `SENTENCE_SEPARATOR`. Default: ` ` (Space/Blank)

Last update: September 14, 2022

Created: September 14, 2022