



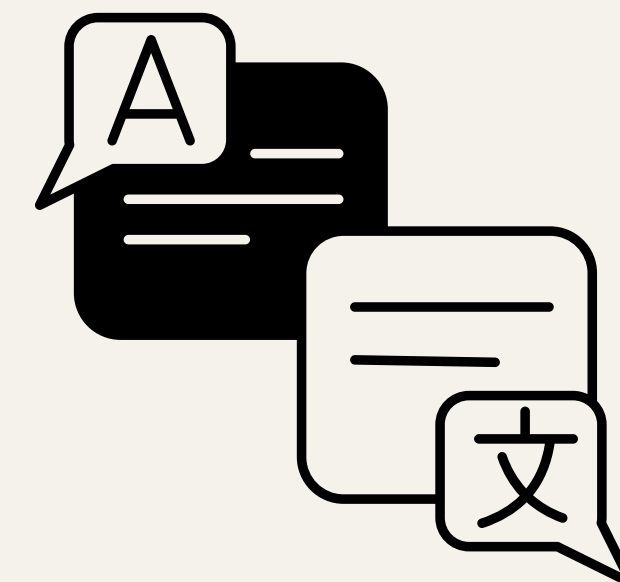
TRABALHO 1

Realizado por:

Alícia Soares Oliveira, PG50169

Ana Carolina Veloso, PG50178

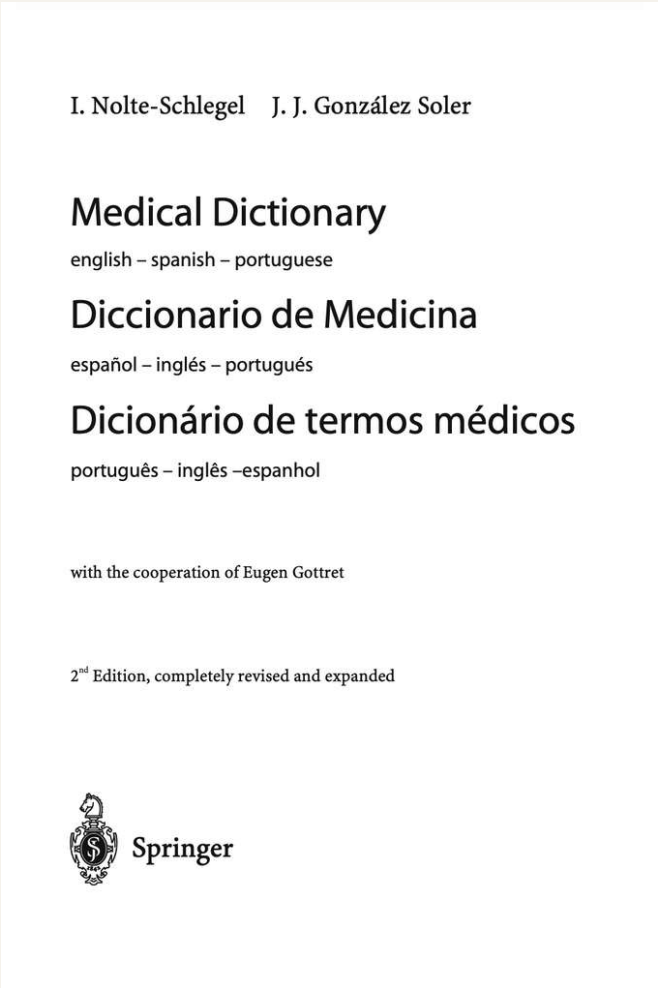
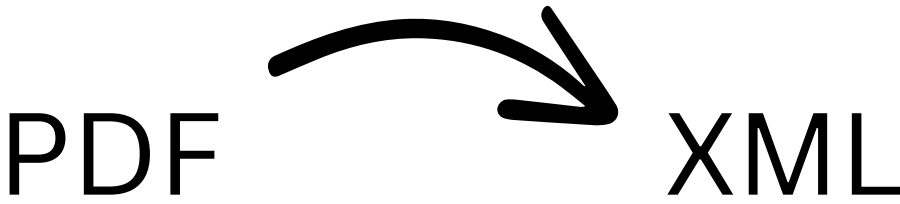
Margarida Ferreira Fernandes, PG50588



Documentos

Selecionaram-se:

- PDF obrigatório
- PDF anatomia geral
- PDF enfermagem



12 Anatomia Geral	
1	ANATOMIA GERAL
2	PARTES DO CORPO HUMANO
3	Cabeça.
4	<i>Fronte.</i> Parte anterior da cabeça. A
5	<i>Occipital.</i> Parte posterior da cabeça (nuca). B
6	<i>Temporo.</i> A
7	<i>Orelha.</i>
8	<i>Face.</i> A
9	<i>Olho.</i>
10	<i>Bochecha.</i> C
11	<i>Nariz.</i>
12	<i>Boca.</i>
13	<i>Mento.</i>
14	<i>Pescoço.</i> Seu limite superior passa por uma linha ao longo da margem inferior da mandíbula, processo mastóide, linha nucal superior, até a protuberância occipital externa; seu limite inferior estende-se da margem superior do mandíbulo do esterno, ao longo da clavícula, ao acrómio e à espinha da escápula, até o processo espinhoso de C VII. A, B
15	Tórax. A, B, C
16	<i>Tórax.</i> Parte do tronco, entre o pescoço e o abdome. Sua estrutura básica é a caixa torácica. Seu limite inferior é a abertura torácica inferior e o diafragma. A
17	<i>Peito.</i> C
18	<i>Abdome.</i> Parte do tronco entre o tórax, a margem superior do sacro, o ligamento inguinal e a sínfise púbica. A, C
19	<i>Pelve.</i> Parte do tronco entre o abdome e o soalho da pelve. A pelve maior e a pelve menor são separadas pela linha terminal.
20	<i>Dorso.</i> Parte posterior do tronco. B
21	<i>Membro superior.</i> Constituído pelo cingulo do membro superior e pela extremidade livre.
22	<i>Cingulo do membro superior.</i> Sua estrutura óssea básica é formada pela escápula e pela clavícula. A, B
23	<i>Axila.</i> Cavidade axilar. Espaço de união entre o membro superior e a parede lateral do tórax. C
24	<i>Braço.</i> A
25	<i>Cotovelo.</i>
26	<i>Antebraço.</i> A
27	<i>Mão.</i>
28	<i>Carpo.</i> A
29	<i>Metacarpo.</i> A
30	<i>Palma.</i> A
31	<i>Dorso da mão.</i>
32	<i>Dedos da mão.</i>
33	<i>Membro inferior.</i> Constituído pelo cingulo do membro inferior e pela extremidade livre.
34	<i>Cingulo do membro inferior.</i> Sua estrutura óssea básica é formada pelo osso do quadril. C
35	<i>Nádegas.</i> B
36	<i>Quadril.</i> Região de união da pelve com a extremidade livre do membro inferior. C
37	<i>Fêmur.</i> A, B
38	<i>Joelho.</i> C
39	<i>Região poplitea.</i> Fossa superficial do joelho. C
40	<i>Perna.</i> A
41	<i>Sura.</i> Panturrilha. B, C
42	<i>Pé.</i>
43	<i>Tarso.</i> A
44	<i>Calcâneo.</i> B
45	<i>Metatarso.</i> A
46	<i>Planta.</i> B
47	<i>Dorso do pé.</i> C
48	<i>Dedos do pé.</i>
49	Cavidades.
50	<i>Cavidade do crânio.</i>
51	<i>Cavidade torácica.</i>
52	<i>Cavidade abdominopélvica.</i> Cavidade conjunta do abdome e da pelve.
53	<i>Cavidade abdominal.</i>
54	<i>Cavidade pélvica.</i>

PDF pt-en-sp

Etapas:

- Limpeza do ficheiro
- Captura da informação relevante
- Organização da informação
- Exposição da informação em JSON

I. Nolte-Schlegel J. J. González Soler

Medical Dictionary

english – spanish – portuguese

Diccionario de Medicina

español – inglés – portugués

Dicionário de termos médicos

português – inglês – espanhol

with the cooperation of Eugen Gottret

2nd Edition, completely revised and expanded



Springer

PDF pt-en-sp

Limpeza

- limpeza das quebras de página;
- limpeza dos <text> e </text>
- limpeza dos itálicos;

r'\n</page>\n<.*>'

```
330 <text top="617" left="517" width="0" height="14" font="14"><b>U</b></text>
331 <text top="609" left="517" width="0" height="14" font="14"><b>inglês</b></text>
332 <text top="582" left="517" width="0" height="14" font="14"><b>U</b></text>
333 <text top="574" left="517" width="0" height="14" font="14"><b>espanhol</b></text>
334 </page>
335 <page number="253" position="absolute" top="0" left="0" height="807" width="540">
336 <text top="76" left="72" width="67" height="11" font="18">bral apoplexy</text>
337 <text top="75" left="145" width="18" height="14" font="17">E</text>
338 <text top="76" left="166" width="47" height="11" font="18">accidente</text>
339 <text top="76" left="215" width="10" height="13" font="19"><i>m</i></text>
```

- limpeza dos vários cabeçalhos e rodapés encontrados no ficheiro;

r'português\n–
\ninglês\n–
\nespanhol\n'

```
259 <b>acidente</b>
260 <b>vascular cerebral (AVC)</b>
261 U
262 cerebrovascular accident, cere-
263 <b>249</b>
264 <b>acidente</b>
265 <b>português</b>
266 <b>U</b>
267 <b>inglês</b>
268 <b>U</b>
269 <b>espanhol</b>
270 bral apoplexy
271 E
272 accidente
273 vas
```


PDF pt-en-sp

Limpeza

- limpeza das palavras divididas por - entre linhas;

`r'\n'`

```
25 abdominal
26 E
27 abdo-
28 minal
29 <b>abdução</b>
30 U
31 abduction
32 E
33 abduc-
34 ción
35 <b>aberração</b>
36 <b>cromossómica</b>
```

- limpeza das páginas pares em que mudava a letra na página seguinte;

`r'.*\n\d{2}[0|2|4|6|8]`
`\n[A-Z]\n'`

```
2591 U
2592 azotemia
2593 E
2594 azotemia
2595 <b>atetose</b>
2596 <b>258</b>
2597 <b>B</b>
2598 <b>bacilo</b>
2599 U
2600 bacillus
2601 E
```

PDF pt-en-sp

Limpeza

- limpeza das letras de mudança de página;
- limpeza de números de página pares e ímpares separadamente;

'.*\n\d{2}
[0|2|4|6|8]\n'

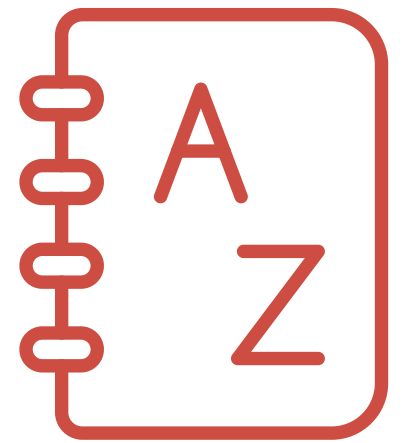
```
489 U
490 get sick
491 E
492 enfermar
493 <b>ácido</b>
494 <b>250</b>
495 <b>adolescência</b>
496 U
```

- limpeza de termos e traduções que estavam em várias linhas, ou seja, tinham \n;

r'\n'

```
552 U
553 aerosol
554 E
555 aerosol
556 <b>afasia</b>
557 <b>(motora/ sensorial/ amnésica/</b>
558 <b>global)</b>
559 U
560 aphasia (motor /sensory/amnesic /global)
561 E
562 afasia
```

PDF pt-en-sp



Captura da informação relevante

- Termos: `r'(.*?)'`
- Inglês: `r'U\n(.*?)\nE'`
- Espanhol: `r'E\n(.*?)\n'`

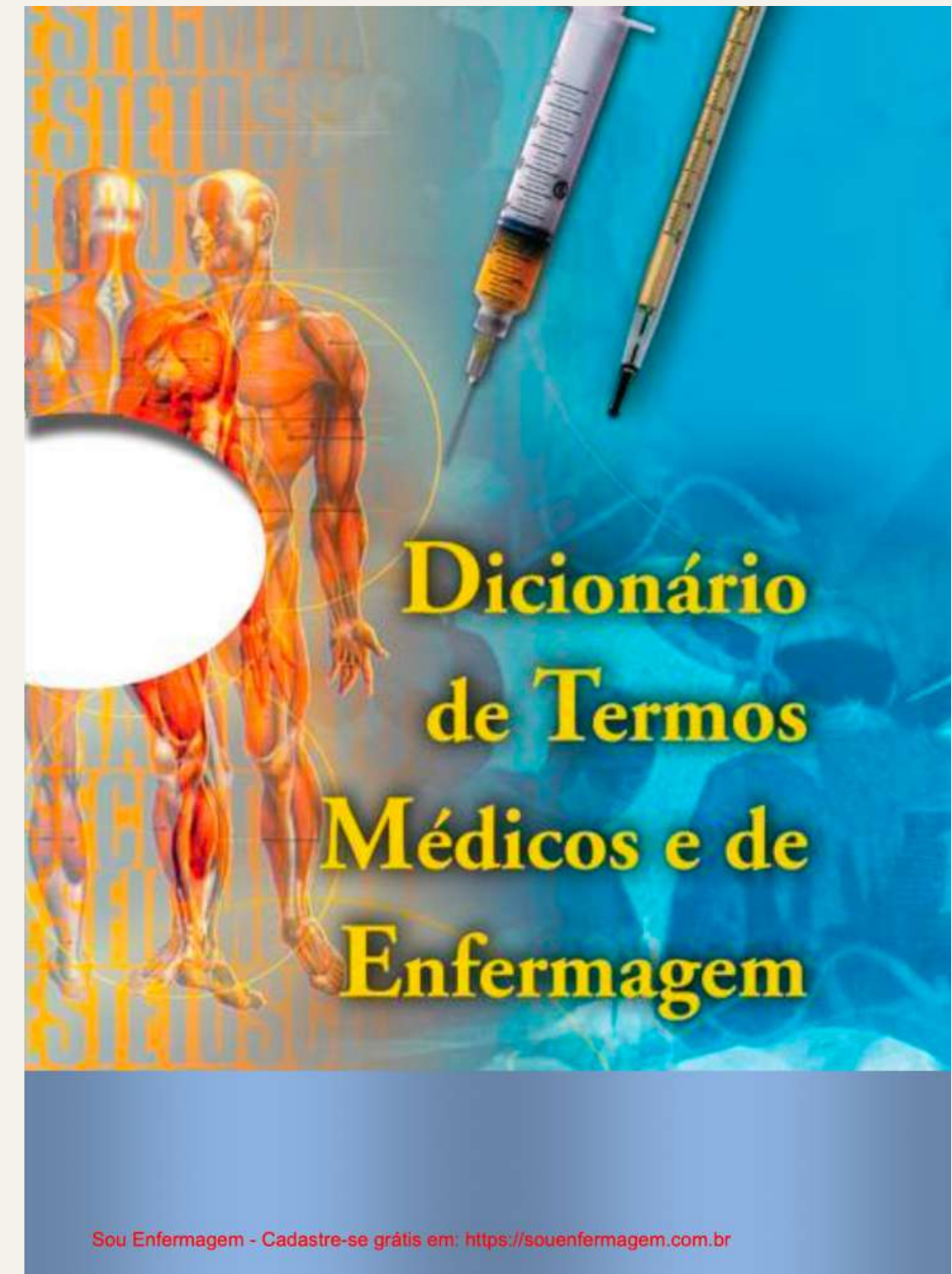
Organização da informação

```
468     "espanhol": "agregación plaquetaria",
469 },
470 "agramatismo": {
471     "ingles": "agrammatism",
472     "espanhol": "agramatismo"
473 },
474 "agranulocitose": {
475     "ingles": "agranulocytosis",
476     "espanhol": "agranulocitosis"
477 },
478 "agravamento": {
479     "ingles": "aggravation, worsening",
480     "espanhol": "agravamiento"
481 },
482 "agravar-se": {
483     "ingles": "get worse, worsen",
484     "espanhol": "agravarse"
485 },
486 "agregação plaquetária": {
487     "ingles": "platelet aggregation",
488     "espanhol": "agregación plaquetar"
489 },
490 "agudilhão": {
```

PDF enfermagem

Etapas:

- Limpeza do ficheiro
- Captura da informação relevante
- Organização da informação
- Exposição da informação em JSON



PDF enfermagem

Limpeza

- limpeza da parte introdutória do ficheiro;
- limpeza dos textos auxiliares próprios de ficheiros XML;

`re.sub(r'</?text.*?>', '', remove_text)`

- limpeza dos vários cabeçalhos e rodapés encontrados no ficheiro;

`re.sub(r'\n[\s*\o\s*]+'+', '\n', remove_text)`

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE pdf2xml SYSTEM "pdf2xml.dtd">

<pdf2xml producer="poppler" version="23.03.0">
<page number="1" position="absolute" top="0" left="0" height="786" width="57
|   <fontspec id="0" size="12" family="Arial" color="#ff0000"/>
<image top="0" left="-1" width="576" height="656" src="Dicionario_de_termos_
<text top="750" left="43" width="391" height="16" font="0">Sou Enfermagem -
```

```
o
o
o
o
o
SHI
SÍF
Sou Enfermagem - Cadastre-se grátis <a href="https://souenfermagem.com.br">em: https:
</page>
<page number="405" position="absolute" top="0" left="0" height="850" width="637">
413
```

PDF enfermagem

Limpeza

- limpeza dos hífens quando há parágrafo nas descrições

```
re.sub(r'\n',"",remove_text)
```

```
lateralmente, pelos músculos abdo-  
minais e, posteriormente, pelas vér-  
tebras e músculos da espinha dorsal.  
Abriga o estômago, os intestinos  
grosso e delgado, o fígado, a vesícu-  
la biliar, o pâncreas, o baço, os rins  
com as correspondentes glândulas  
supra-renais, a aorta abdominal, va-
```

- limpeza dos <i>'s e dos hífens entre <i>'s, mantendo apenas a palavra, juntandos os 2 <i>'s
- limpeza dos <i> mantendo apenas as palavras no seu interior

```
re.sub(r'\n?<i>(.*?)</i>\n', r'\1', remove_text)
```

```
<b>AÇÚCAR DE FRUTA</b>  
V.  
<i>Levu-</i>  
<i>lose</i>  
.
```

PDF enfermagem

Limpeza

- limpeza das capas iniciais de cada letra;

```
re.sub(r'([A-Z])\1<b>', r'<b>', remove_text)
```

coração que faz os sangues ve
| BBB.C.G.

- limpeza da parte final do ficheiro;

```
re.sub(r'<item .*\\n(.*)\\n*</pdf2xml>', ' ', remove_text)
```

```
<item page="1">Capa</item>
<item page="2">Expediente</item>
<item page="3">Introdução</item>
<item page="15">A</item>
<item page="69">B</item>
<item page="85">C</item>
<item page="139">D</item>
<item page="171">E</item>
<item page="206">F</item>
<item page="224">G</item>
<item page="239">H</item>
<item page="260">I</item>
<item page="275">J K L</item>
<item page="291">M</item>
<item page="313">N O</item>
<item page="335">P</item>
<item page="378">Q R</item>
<item page="397">S</item>
<item page="417">T</item>
<item page="441">U V Z</item>
</outline>
</pdf2xml>
```

PDF enfermagem

Limpeza

- limpeza dos artefactos vários encontrados nos bolds (), de forma uniformizar a informação entre ;
- limpeza e uniformização das descrições, ou seja, verificou-se que a ultima descrição nao encaixava no pattern das restantes.

```
re.sub(r'<b>\s</b>', r' ', remove_text)
```

```
<b>fi</b>  
<b> cam frequentemente surpresos </b>  
<b>ao saber que seus testes de fígado dão anormal  
<b>de rotina.</b>
```

```
<b> . </b>
```

```
<b>1 Visão normal</b>  
- A luz de um objeto externo é trazida  
para um foco na retina sensível do  
fundo do globo ocular.  
<b>2 Hipermetropia </b>
```

```
<b> , </b>
```

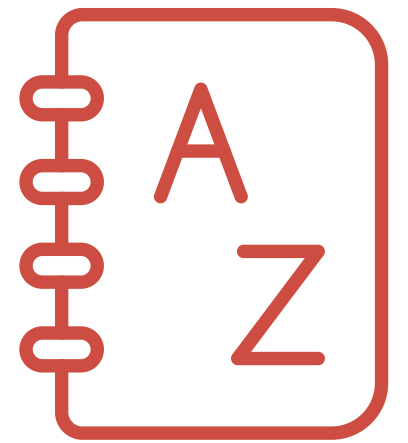
```
V. Ácido. <b>ACRO</b>
```

```
<b>s</b>  
êmen
```

```
re.sub(r'Vertigens\\.\\)', r'Vertigens.)\n<b>', remove_text)
```

```
de Menière". (V. Surdez, Vertigens.)
```


PDF enfermagem



Captura da informação relevante

Organização da informação

```
re.findall(r'<b>(.*?)</b>\s*(.*?)\s*(?=<b>|$)', remove_text, re.DOTALL)
```

- Termos:
 - identificados por estarem rodeados por
- Descrições
 - identificados por estarem rodeados por e



KEY



VALUE

PDF enfermagem

Escrita em JSON

```
"ABERRAÇÃO": {  
  "designacao": "Desvio do normal. Genética Anomalia na situação ou na conformação de um órgão",  
},  
"ABERRAÇÕES CROMOSSÔMICAS": {  
  "designacao": "Alteração na anatomia dos cromossomos normais que geralmente afetam a função",  
},  
"ABERRANTE": {  
  "designacao": "Que se desvia do normal, do padrão comum. Ex.: artéria aberrante, veia aberrante",  
},  
"ABLAÇÃO": {  
  "designacao": "Separação por incisão ou amputação cirúrgica de qualquer parte do corpo, por exemplo",  
},  
"ABLEPSIA": {  
  "designacao": "Cegueira, perda ou falta de visão.",  
},  
"ABLUÇÃO": {  
  "designacao": "Banho, lavagem. Ato de lavar-se, banhar-se.",  
},  
"ABORTAMENTO": {  
  "designacao": "Expulsão do feto antes de 180 dias de gestação. Depois desse prazo, chama-se parto",  
},
```

PDF anatomia geral

Etapas:

- Limpeza do ficheiro
- Captura da informação relevante
- Organização da informação
- Exposição da informação em JSON

12 Anatomia Geral	
1	1 ANATOMIA GERAL
	2 PARTES DO CORPO HUMANO
	3 Cabeça.
	4 Fronte. Parte anterior da cabeça. A
	5 Occipital. Parte posterior da cabeça (nuca). B
	6 Têmpora. A
	7 Orelha.
	8 Face. A
	9 Olho.
	10 Bochecha. C
	11 Nariz.
	12 Boca.
	13 Mento.
	14 Pescoço. Seu llmite superior passa por uma linha ao longo da margem inferior da mandíbula, processo mastóide, linha nugal superior, até a protuberância occipital externa; seu limite inferior estende-se da margem superior do manúbrio do esterno, ao longo da clavícula, ao acrômio e à espinha da escápula, até o processo espinhoso de C VII. A, B
	15 Tronco. A, B, C
	16 Tórax. Parte do tronco, entre o pescoço e o abdome. Sua estrutura básica é a caixa torácica. Seu limite inferior é a abertura torácica inferior e o diafragma. A
	17 Peito. C
	18 Abdome. Parte do tronco entre o tórax, a margem superior do sacro, o ligamento inguinal e a sínfise púbica. A, C
	19 Pelve. Parte do tronco entre o abdome e o soalho da pelve. A pelve maior e a pelve menor são separadas pela linha terminal.
	20 Dorso. Parte posterior do tronco. B
	21 Membro superior. Constituído pelo cingulo do membro superior e pela extremidade livre.
	22 Cingulo do membro superior. Sua estrutura óssea básica é formada pela escápula e pela clavícula. A, B
	23 Axila. Cavidade axilar. Espaço de união entre o membro superior e a parede lateral do tórax. C
	24 Braço. A
	25 Cotovelo.
	26 Antebraço. A
	27 Mão.
	28 Carpo. A
	29 Metacarpo. A
	30 Palma. A
	31 Dorso da mão.
	32 Dedos da mão.
	33 Membro inferior. Constituído pelo cingulo do membro inferior e pela extremidade livre.
	34 Cingulo do membro inferior. Sua estrutura óssea básica é formada pelo osso do quadril. C
	35 Nádegas. B
	36 Quadril. Região de união da pelve com a extremidade livre do membro inferior. C
	37 Fêmur. A, B
	38 Joelho. C
	39 Região poplítea. Fossa superficial do joelho. C
	40 Perna. A
	41 Sura. Panturrilha. B, C
	42 Pé
	43 Tarso. A
	44 Calcanhar. B
	45 Metatarso. A
	46 Planta. B
	47 Dorso do pé. C
	48 Dedos do pé.
	49 Cavidades.
	50 Cavidade do crânio.
	51 Cavidade torácica.
	52 Cavidade abdominopélvica. Cavidade conjunta do abdome e da pelve.
	53 Cavidade abdominal.
	54 Cavidade pélvica.

PDF anatomia geral - Etapa 1 - limpeza

Remoção de marcas:

```
text = re.sub(r"</?page.*>", "", text)
```

```
text = re.sub(r"</?image.*?>", "", text)
```

```
text = re.sub(r"</?fontspec.*?>", "", text)
```

```
text = re.sub(r"<[?]xml(.*)?>", "", text)
```

```
text = re.sub(r"<[!]DOCTYPE(.*)?>", "", text)
```

```
text = re.sub(r"</?pdf(.*)?>", "", text)
```

```
teste_anatomia.xml
1  <?xml version="1.0" encoding="UTF-8"?>
2  <!DOCTYPE pdf2xml SYSTEM "pdf2xml.dtd">
3
4  <pdf2xml producer="poppler" version="22.08.0">
5  <page number="1" position="absolute" top="0"
6  left="0" height="807" width="552">
7      <fontspec id="0" size="11" family="JJHGOL
8      +DTLArgoTBlack" color="#231f20"/>
9      <fontspec id="1" size="11" family="JJHGOM
10     +DTLArgoT" color="#146eaf"/>
11     <fontspec id="2" size="12" family="JJHGOL
12     +DTLArgoTBlack" color="#ffffff"/>
13     <fontspec id="3" size="10" family="JJHHDA
14     +CaxtonBT-Light" color="#231f20"/>
15     <fontspec id="4" size="9" family="JJHHDB
16     +DTLArgoT-Bold" color="#146eaf"/>
17     <fontspec id="5" size="9" family="JJHGOM
18     +DTLArgoT" color="#231f20"/>
19     <fontspec id="6" size="9" family="JJHHFD
20     +DTLArgoTMedium" color="#231f20"/>
21     <fontspec id="7" size="9" family="JJHHFE
22     +DTLArgoTLight" color="#231f20"/>
23     <text top="57" left="55" width="12" height="11"
24     font="0">12</text>
25     <text top="57" left="68" width="84" height="11"
26     font="1"><b> Anatomia </b></text>
27     <text top="57" left="143" width="27" height="11"
28     font="1"><b>Geral</b></text>
29     <text top="98" left="29" width="7" height="12"
```


PDF anatomia geral - Etapa 1 - limpeza

Remoção de frases inicializadas com *, letras isoladas s/espacos, letras isoladas e combinadas c/espacos, marcas de páginas:

```
text = re.sub(r"<text(.*)>\s*(\s*)(.*)</text>", "", text)
```

```
text = re.sub(r'<text[^>]*><b>[A-Z]</b>\s*</text>\n', "", text)
```

```
text = re.sub(r'<text[^>]*>\s*(<b>)?[ABC](,\s*[ABC])*(</b>)?\s*</text>', "", text)
```

```
text = re.sub(r"A, F", "", text)
```

```
text = re.sub(r"C VII", "", text)
```

```
text = re.sub(r"Ver pág. 40", "", text)
```

```
text = re.sub(r"\s+", "", text)
```

```
nt="7"><i> A, B
ont="3">
nt="5"><b> Pescoço.</b></text>
font="3"> Seu limite superior passa por uma </text>
font="3">linha ao longo da margem inferior da mandíbula,
font="3">bula, processo mastóide, linha nugal superior, </text>
font="3">até a protuberância occipital externa; seu limite inferior
font="3">estende-se da margem superior do </text>
font="3">manúbrio do esterno, ao longo da clavícula, ao </text>
font="3">acrômio e à espinha da escápula, até o processo </text>
ont="3">espinhoso de C VII. A, B</text>
ont="3"> 15</text>
nt="5"><b> Tronco.</b></text>
font="3"> A, B, C</text>
ont="3"> 16</text>
nt="6"><i> Tórax.</i></text>
font="3"> Parte do tronco, entre o pescoço e o abdome. Sua estrutura
font="3">básica é a caixa torácica. Seu limite inferior é a abertura
font="3">torácica inferior e o diafragma. A</text>
ont="3"> 17</text>
nt="7"><i> Peito.</i></text>
ont="3"> C</text>
ont="3"> 18</text>
nt="6"><i> Abdome.</i></text>
font="3"> Parte do tronco entre o tórax, a margem superior do sacro,
font="3">o ligamento inguinal e a sínfise púbica. A, C</text>
ont="3"> 19</text>
```

PDF anatomia geral - Etapa 1 - limpeza

Substituir linhas dos números pelo marcador #, delimitar expressão com , re-estruturar expressões reiteradas

```
text = re.sub(r'<text.*?>\s*\d+(\.?)+(\.\\d+)?([;,]?)\s*\d+(\.\\d+)?)*\s*</text>', '#', text)
```

```
text = re.sub(r"#\n<text(.*)>(.)</text>", r"<text\1><b>\2</b></text>", text)
```

```
<text top="57" left="55" width="12" height="11" font="0">12</text>
<text top="57" left="68" width="84" height="11" font="1"><b> Anatomia <
<text top="57" left="143" width="27" height="11" font="1"><b>Geral</b></
<text top="98" left="29" width="7" height="12" font="2">1</text>
<text top="94" left="55" width="13" height="10" font="3"> 1</text>
<text top="94" left="68" width="73" height="9" font="4"><i><b> ANATOMIA
<text top="106" left="55" width="13" height="10" font="3"> 2</text>
<text top="106" left="68" width="114" height="9" font="4"><i><b> PARTES
<text top="119" left="55" width="13" height="10" font="3"> 3</text>
<text top="119" left="68" width="36" height="9" font="5"><b> Cabeça.</b>
<text top="131" left="55" width="13" height="10" font="3"> 4</text>
<text top="132" left="68" width="31" height="9" font="6"><i> Fronte.</i>
<text top="131" left="99" width="114" height="10" font="3"> Parte anter
<text top="144" left="55" width="13" height="10" font="3"> 5</text>
<text top="144" left="68" width="40" height="9" font="6"><i> Occipital.<
<text top="144" left="108" width="146" height="10" font="3"> Parte post
<text top="156" left="55" width="13" height="10" font="3"> 6</text>
<text top="157" left="68" width="41" height="9" font="6"><i> Têmpora.</i>

<text top="169" left="55" width="13" height="10" font="3"> 7</text>
<text top="169" left="68" width="32" height="9" font="6"><i> Orelha.</i>
<text top="169" left="100" width="3" height="10" font="3"> </text>
<text top="181" left="55" width="13" height="10" font="3"> 8</text>
<text top="182" left="68" width="24" height="9" font="6"><i> Face.</i></
```


PDF anatomia geral - Etapa 1 - limpeza

text = re.sub(r">", ">", text)

```
"84" height="11" font="1"><b> Anatomia </b></b></text>
="27" height="11" font="1"><b>Geral</b></text>

"73" height="9" font="4"><i><b> ANATOMIA GERAL</b></i><b></b></text>
="114" height="9" font="4"><i><b> PARTES DO CORPO HUMANO</b></i><b></b></text>
="36" height="9" font="5"><b> Cabeça.</b><b></b></text>
="31" height="9" font="6"><i> Fronte.</i><b></b></text>
="114" height="10" font="3"> Parte anterior da cabeça. A</text>
="40" height="9" font="6"><i> Occipital.</i><b></b></text>
h="146" height="10" font="3"> Parte posterior da cabeça (nuca). B</text>
="41" height="9" font="6"><i> Têmpora.</i><b></b></text>

="32" height="9" font="6"><i> Orelha.</i><b></b></text>
h="3" height="10" font="3"> </text>
="24" height="9" font="6"><i> Face.</i><b></b></text>

="24" height="9" font="7"><i> Olho.</i><b></b></text>
="42" height="9" font="7"><i> Bochecha.</i><b></b></text>

="26" height="9" font="7"><i> Nariz.</i><b></b></text>
="25" height="9" font="7"><i> Boca.</i><b></b></text>
="30" height="9" font="7"><i> Mento.</i><b></b></text>
="41" height="9" font="5"><b> Pescoço.</b><b></b></text>
h="163" height="10" font="3"> Seu limite superior passa por uma </text>
```

text = re.sub(r"#+(.*)", "", text)

```
ft="68" width="84" height="11" font="1"><b> Anatomia </b></text>
ft="143" width="27" height="11" font="1"><b>Geral</b></text>

ft="68" width="73" height="9" font="4"><i><b> ANATOMIA GERAL</b></i></text>
eft="68" width="114" height="9" font="4"><i><b> PARTES DO CORPO HUMANO</b></i></text>
eft="68" width="36" height="9" font="5"><b> Cabeça.</b></text>
eft="68" width="31" height="9" font="6"><i> Fronte.</i></text>
eft="99" width="114" height="10" font="3"> Parte anterior da cabeça. A</text>
eft="68" width="40" height="9" font="6"><i> Occipital.</i></text>
eft="108" width="146" height="10" font="3"> Parte posterior da cabeça (nuca). B</text>
eft="68" width="41" height="9" font="6"><i> Têmpora.</i></text>

eft="68" width="32" height="9" font="6"><i> Orelha.</i></text>
eft="100" width="3" height="10" font="3"> </text>
eft="68" width="24" height="9" font="6"><i> Face.</i></text>

eft="68" width="24" height="9" font="7"><i> Olho.</i></text>
eft="68" width="42" height="9" font="7"><i> Bochecha.</i></text>

eft="68" width="26" height="9" font="7"><i> Nariz.</i></text>
eft="68" width="25" height="9" font="7"><i> Boca.</i></text>
eft="68" width="30" height="9" font="7"><i> Mento.</i></text>
eft="68" width="41" height="9" font="5"><b> Pescoço.</b></text>
eft="109" width="163" height="10" font="3"> Seu limite superior passa por uma </text>
```

PDF anatomia geral - Etapa 1 - limpeza

Juntar as expressões que estão separadas pelas regras de translineação, substituir marcadores, remoção <text> e </text>

```
text = re.sub(r"-</b></text>\n<text(.*)><b>", "", text)
```

```
text = re.sub(r"\[\\(.*)\\]", r"\1", text)
```

```
text = re.sub(r"<i><b>(.)</b></i>([^\<]*)", r"<t>\1</t>\2", text)
```

```
text = re.sub(r"<i>(.)</i>([^\<]*)", r"<b>\1</b>\2", text)
```

```
text = re.sub(r"</?text.*?>", "", text)
```

```
text = re.sub(r"-\\n", "", text)
```

```
<text top="702" left="72" width="29" height="10" font="3">coço. B</text>
<text top="715" left="68" width="200" height="9" font="5"><b> Região cervical anterior (trígono cervical ante</b></
text>
<text top="726" left="72" width="21" height="9" font="5"><b>rior).</b></text>
<text top="726" left="93" width="175" height="10" font="3"> Trígono entre a linha mediana do pesco</text>
<text top="94" left="302" width="196" height="10" font="3">co - a margem anterior do m. esternocleidomas</text>
```

```
= "10" font="3">glândula parótida e o m. masseter. B</text>
= "9" font="6"><i> [[Fossa retromandibular]].</i></text>
= "10" font="3"> Ao longo do ramo da </text>
```

```
= "4"><t> ANATOMIA GERAL</t></text>
nt="4"><t> PARTES DO CORPO HUMANO</t></text>
t="5"><b> Cabeça.</b></text>
```

```
<b> Anatomia </b>
<b>Geral</b>
```

```
<t> ANATOMIA GERAL</t>
<t> PARTES DO CORPO HUMANO</t>
<b> Cabeça.</b>
```


PDF anatomia geral - Etapa 1 - limpeza

Remover marcadores, remover numeração romana vinculada a imagens, substituição de espaços

```
text = re.sub(r"\bm+\.", "", text)
```

```
text = re.sub(r"<b> </b>", "", text)
```

```
text = re.sub(r"<b>\.</b>", "", text)
```

```
text = re.sub(r"<b>(I+)</b>", r"\I", text)
```

```
text = re.sub(r"\bIII\b\n\bII\b\n\bI\b\n\bI\b\n\bV\b", "", text)
```

```
text = re.sub(r"\n(\s*)\n", "\n", text)
```

```
<b> Anatomia </b>
<b>Geral</b>
<t> ANATOMIA GERAL</t>
<t> PARTES DO CORPO HUMANO</t>
<b> Cabeça.</b>
<b> Fronte.</b>
| Parte anterior da cabeça. A
<b> Occipital.</b>
| Parte posterior da cabeça (nuca). B
<b> Têmpora.</b>
<b> Orelha.</b>
<b> Face.</b>
<b> Olho.</b>
<b> Bochecha.</b>
```

PDF anatomia geral - Etapas 2 e 3

captura e organização



Captura da informação
relevante

Organização da
informação

```
list = re.findall(r"<b>(.*?)</b>([^\<]*)", text)
```

- Termos:
 - identificados por estarem rodeados por
- Descrições
 - identificados por estarem rodeados por e



KEY



VALUE

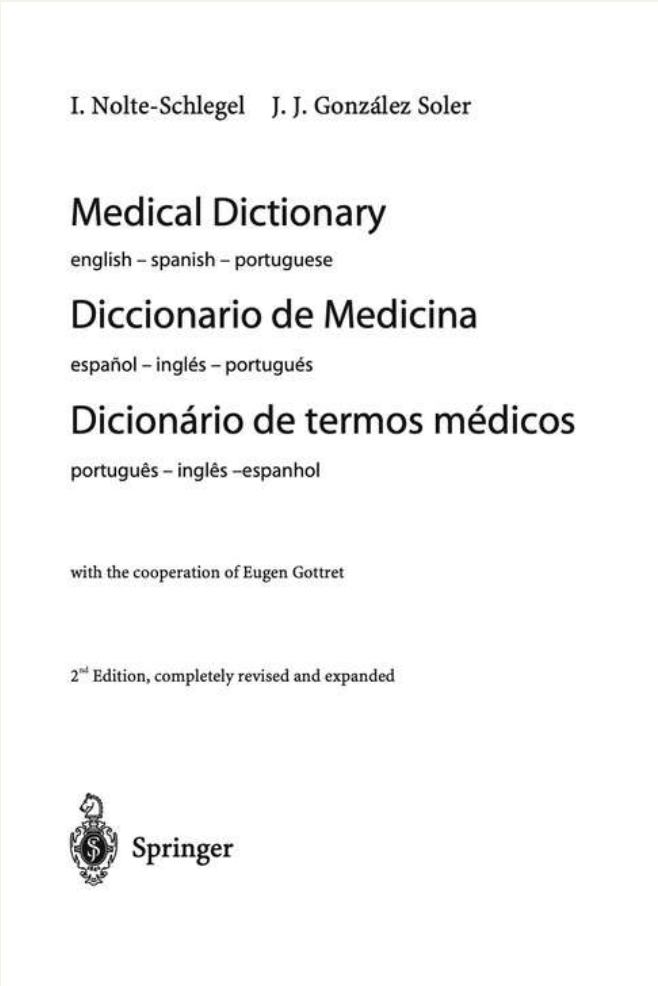
PDF anatomia geral - Etapa 4

json final

```
,
"intermédio": {
  "descricao": "Situado entre duas estruturas"
},
"medial": {
  "descricao": "Situado próximo do plano mediano"
},
"lateral": {
  "descricao": "Situado afastado do plano mediano"
},
"anterior": {
  "descricao": "Situado à frente de"
},
"posterior": {
  "descricao": "Situado atrás de"
},
"ventral": {
  "descricao": "Situado em direção ao ventre"
},
"dorsal": {
  "descricao": "Situado em direção ao dorso"
},
"frontal": {
  "descricao": "Pertencente à frente (testa); situado paralelamente à frente (testa)"
},
"superior": {
  "descricao": "Situado em direção à cabeça"
},
"inferior": {
  "descricao": "Situado em direção à base"
}
```

Resultado Final: junção dos 3 ficheiros json

- PDF obrigatório
- PDF anatomia geral
- PDF enfermagem



12 Anatomia Geral	
1	ANATOMIA GERAL
2	PARTES DO CORPO HUMANO
3	Cabeça.
4	<i>Fronte.</i> Parte anterior da cabeça. A
5	<i>Occipital.</i> Parte posterior da cabeça (nuca). B
6	<i>Temporo.</i> A
7	<i>Orelha.</i>
8	<i>Face.</i> A
9	<i>Olho.</i>
10	<i>Bochecha.</i> C
11	<i>Nariz.</i>
12	<i>Boca.</i>
13	<i>Mento.</i>
14	<i>Pescoço.</i> Seu limite superior passa por uma linha ao longo da margem inferior da mandíbula, processo mastóide, linha nucal superior, até a protuberância occipital externa; seu limite inferior estende-se da margem superior do mandíbulo do esterno, ao longo da clavícula, ao acrómio e à espinha da escápula, até o processo espinhoso de C VII. A, B
15	Tórax. A, B, C
16	<i>Tórax.</i> Parte do tronco, entre o pescoço e o abdome. Sua estrutura básica é a caixa torácica. Seu limite inferior é a abertura torácica inferior e o diafragma. A
17	<i>Peito.</i> C
18	<i>Abdome.</i> Parte do tronco entre o tórax, a margem superior do sacro, o ligamento inguinal e a sínfise púbica. A, C
19	<i>Pelve.</i> Parte do tronco entre o abdome e o soalho da pelve. A pelve maior e a pelve menor são separadas pela linha terminal.
20	<i>Dorso.</i> Parte posterior do tronco. B
21	<i>Membro superior.</i> Constituído pelo cingulo do membro superior e pela extremidade livre.
22	<i>Cingulo do membro superior.</i> Sua estrutura óssea básica é formada pela escápula e pela clavícula. A, B
23	<i>Axila.</i> Cavidade axilar. Espaço de união entre o membro superior e a parede lateral do tórax. C
24	<i>Braço.</i> A
25	<i>Cotovelo.</i>
26	<i>Antebraço.</i> A
27	<i>Mão.</i>
28	<i>Carpo.</i> A
29	<i>Metacarpo.</i> A
30	<i>Palma.</i> A
31	<i>Dorso da mão.</i>
32	<i>Dedos da mão.</i>
33	<i>Membro inferior.</i> Constituído pelo cingulo do membro inferior e pela extremidade livre.
34	<i>Cingulo do membro inferior.</i> Sua estrutura óssea básica é formada pelo osso do quadril. C
35	<i>Nádegas.</i> B
36	<i>Quadril.</i> Região de união da pelve com a extremidade livre do membro inferior. C
37	<i>Fêmur.</i> A, B
38	<i>Joelho.</i> C
39	<i>Região poplíteo.</i> Fossa superficial do joelho. C
40	<i>Perna.</i> A
41	<i>Sura.</i> Panturrilha. B, C
42	<i>Pé.</i>
43	<i>Tarso.</i> A
44	<i>Calcâneo.</i> B
45	<i>Metatarso.</i> A
46	<i>Planta.</i> B
47	<i>Dorso do pé.</i> C
48	<i>Dedos do pé.</i>
49	Cavidades.
50	<i>Cavidade do crânio.</i>
51	<i>Cavidade torácica.</i>
52	<i>Cavidade abdominopélvica.</i> Cavidade conjunta do abdome e da pelve.
53	<i>Cavidade abdominal.</i>
54	<i>Cavidade pélvica.</i>

Resultado Final: junção dos 3 ficheiros json

```
"sinartrose": {
  "designacao": "Articulação que não tem nenhum movimento. Exemplo: a dos ossos do crânio.",
  "ingles": "synarthrosis, synarthrodial joint",
  "espanhol": "sinartrosis",
  "descricao": "Ligação de ossos por meio de tecido conectivo fi broso denso ou cartilagem"
},
"arteríola": {
  "designacao": "Pequena artéria.",
  "ingles": "arteriole",
  "espanhol": "arteriola",
  "descricao": "Pequena artéria diretamente antes do capilar"
},
"indicador": {
  "designacao": "Substância que muda de cor para indicar determinada reação.",
  "ingles": "index (finger)",
  "espanhol": "indicador",
  "descricao": "O segundo dedo da mão (II)"
},
"olho": {
  "descricao": "",
  "ingles": "eye",
  "espanhol": "ojo"
},
"caudal": {
  "designacao": "Referente à cauda.",
  "ingles": "caudal",
```

Dicionário Multilíngue - Multilingual Dictionary - Diccionario Multilingüe

Este é um dicionário de termos médicos que inclui 3 línguas: português, inglês e espanhol.

Alguns termos contêm ainda a descrição/designação associada.

This is a dictionary of medical terms that includes 3 languages: portuguese, english and spanish.

Some terms also contain the associated description/designation.

Este es un diccionario de términos médicos que incluye 3 idiomas: portugués, inglés y español.

Algunos términos también contienen la descripción/designación asociada.

Com a contínua evolução e a complexidade dos termos da área da saúde, estudantes, enfermeiros, técnicos de enfermagem e outros profissionais da área precisam de uma obra de referência rápida, disponível, compacta e completa. Com o objetivo de esclarecer e simplificar a linguagem das ciências da saúde, elaboramos o Dicionário de Termos Médicos e de Enfermagem. Além da explicação de milhares de termos, este dicionário inclui uma introdução sobre alimentação, hábitos saudáveis e outros tópicos importantes para profissionais que valorizam a qualidade de vida e investem na excelência do seu trabalho.

Conteúdo dos Dicionários Médicos © Copyright – todos os direitos reservados: [Editora Rideel](#)