

Correção: Campo 'Abstract'

01. DADOS GERAIS

Linguagem de Programação:	Scala
Build:	SBT (Scala Build Tool)
Categoria:	Correção

02. ERRO/DEFEITO

1. O Valor da Chave 'Abstract' nos documentos gerados no mongoDB pelo projeto SparkCovid estão com o seu valor preenchido em alguns documentos com quebra de linha e ponto e vírgula (;) no lugar do pipe, conforme imagem abaixo:

```
id: ObjectId('646d61cc9f8c5d467c7620c2')
Abstract: "To the Editor:
          In the battle against COVID-19, scientists all over the..."
Accession Number: ""
AlternateId: ""
Authors: "Li, H.| Xu, J.| Cao, B."
CovNum: "1578734"
Database: "Embase"
Date Added: ""
DOI: "10.1016/j.chest.2021.02.075"
FulltextLink: "https://doi.org/10.1016/j.chest.2021.02.075"
Issue: "1"
Journal: "Chest"
Keywords: "ChiCTR2000029308|lopinavir|virus, RNA|anal, swab|clinical, outcome|cor..."
KJD: ""
Language: "en"
Pages: "e86"
PMID: ""
PublishDate: ""
Published Month: ""
PublishedYear: "2021"
SCIELO: ""
Tags: ""
Title: "Response"
UNKNOWN: ""
Volume: "160"
WOS: ""
_updd: 2023-05-23T03:00:00.000+00:00
_upddSrc: 2022-09-15T03:00:00.000+00:00
```

2. A incorreta definição do valor foi encontrada através do modulo 'ScalaMongoDiff' comparando duas coleções no mongoDB. Uma coleção incluída pelo SparkCovid e outra pelo MongoDBMigrations lendo diretamente as planilhas do PENTAH0.

```
_id: ObjectId('646d85917fec5536e588bd2e')
CovNum: "1578734"
~ Abstract: Array
  0: "To the Editor: In the battle against COVID-19, scientists all over the..."
  1: "To the Editor: In the battle against COVID-19, scientists all over the..."
  2: "Diff: In the battle against COVID-19, scientists all over the world a..."
~ Date Added: Array
  0: ""
  1: "16/09/2022"
  2: "Diff: 16/09/2022"
_updd: "Wed May 24 00:33:36 BRT 2023"
```

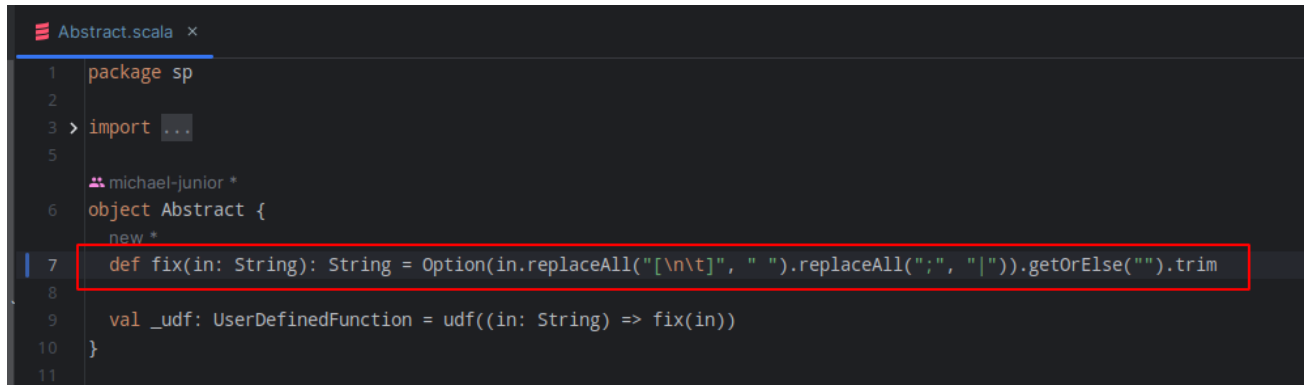
03. ESPERADO

1. É esperado que o valor da chave 'Abstract' tenha o conteúdo sem quebra de linhas ou tab, conforme a imagem abaixo:

```
_id: ObjectId('646e1ff7c1229e7224f85395')
Abstract: "To the Editor: In the battle against COVID-19, scientists all over the..."
Accession Number: ""
AlternateId: ""
Authors: "Li, H.| Xu, J.| Cao, B."
CovNum: "1578734"
Database: "Embase"
Date Added: ""
DOI: "10.1016/j.chest.2021.02.075"
FulltextLink: "https://doi.org/10.1016/j.chest.2021.02.075"
Issue: "1"
Journal: "Chest"
Keywords: "ChiCTR2000029308|lopinavir|virus, RNA|anal, swab|clinical, outcome|cor..."
KJD: ""
Language: "en"
Pages: "e86"
PMID: ""
PublishDate: ""
Published Month: ""
Published Year: "2021"
SCIELO: ""
Tags: ""
Title: "Response"
UNKNOWN: ""
Volume: "160"
WOS: ""
_updd: 2023-05-24T03:00:00.000+00:00
_upddSrc: 2022-09-15T03:00:00.000+00:00
```

04. CORREÇÃO NO CÓDIGO

1. No objeto 'Abstract' deve se aplicar um `replaceAll` para substituir a quebra de linha e tab por um espaço e ponto e vírgula com pipe.



```
Abstract.scala x
1 package sp
2
3 > import ...
4
5
6 object Abstract {
7   new *
8   def fix(in: String): String = Option(in.replaceAll("[\\n\\t]", " ").replaceAll(";", "|")).getOrElse("").trim
9
10   val _udf: UserDefinedFunction = udf((in: String) => fix(in))
11 }
```