

Correção: Campo 'FulltextLink'

01. DADOS GERAIS

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

02. ERRO/DEFEITO

1. O Valor da Chave 'FulltextLink' nos documentos gerados no mongoDB pelo projeto SparkCovid estão com o seu valor preenchido sempre com 'https://pesquisa.bvsalud.org/portal/resource/en/mdl-' quando deveria ser vazio, conforme imagem abaixo:

```
_id: ObjectId('646d11f9166b4a4e58e2c423')
Abstract: "A 53-year-old man developed severe thrombocytopenia during off-label t..."
Accession Number: ""
AlternateId: ""
Authors: ""
CovNum: "1600165"
Database: "EuropePMC"
Date Added: ""
DOI: ""
FulltextLink: "https://pesquisa.bvsalud.org/portal/resource/en/mdl-"
Issue: "1"
Journal: "Reactions Weekly"
Keywords: ""
KJD: ""
Language: "en"
Pages: "190-190"
PMID: ""
PublishDate: ""
Published Month: ""
Published Year: "2021"
SCIELO: ""
Tags: ""
Title: "Favipiravir/levofloxacin/paracetamol : Severe thrombocytopenia and off..."
UNKNOWN: ""
Volume: "1886"
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 PENTAHO.

```
_id: ObjectId('646d85917fec5536e588bd30')
CovNum: "1600165"
FulltextLink: Array
  0: "https://pesquisa.bvsalud.org/portal/resource/en/mdl-"
  1: ""
  2: "Diff: "
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 'FulltextLink' tenha o conteúdo vazio sempre que as chaves 'DOI' e 'PMID' forem vazios, conforme a imagem abaixo:

```
_id: ObjectId('646dfcd17aade637bab3d418')
Abstract: "A 53-year-old man developed severe thrombocytopenia during off-label t..."
Accession Number: ""
AlternateId: ""
Authors: ""
CovNum: "1600165"
Database: "EuropePMC"
Date Added: ""
DOI: ""
FulltextLink: ""
Issue: "1"
Journal: "Reactions Weekly"
Keywords: ""
KJD: ""
Language: "en"
Pages: "190-190"
PMID: ""
PublishDate: ""
Published Month: ""
Published Year: "2021"
SCIELO: ""
Tags: ""
Title: "Favipiravir/levofloxacin/paracetamol : Severe thrombocytopenia and off..."
UNKNOWN: ""
Volume: "1886"
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 'FulltextLink' deve se corrigir as condições para: Se flt e PMID estão sendo vazios o conteúdo de 'FulltextLink' deve ser vazio.

```
FulltextLink.scala x
1 package sp
2
3 import org.apache.spark.sql.expressions.UserDefinedFunction
4 import org.apache.spark.sql.functions.udf
5
6 michael-junior
7 object FulltextLink {
8   michael-junior
9   def fix(doi: String,
10         pmid:String): String = {
11     val DOI = Option(doi).getOrElse("").trim
12     val flt = DOI.replace("DOI:", "").trim
13     val PMID = Option(pmid).getOrElse("").trim
14
15     if (flt.isEmpty) {
16       if (PMID.isEmpty) "https://pesquisa.bvsalud.org/portal/resource/en/mdl-" + PMID
17       else ""
18     } else if (flt.contains("http")) flt
19     else if (flt.contains("doi.org")) "https://" + DOI
20     else "https://doi.org/" + DOI
21   }
22
23   val _udf: UserDefinedFunction = udf((doi: String, pmid: String) => fix(doi, pmid))
24 }
```