

Extração de Dados da Rede Social Twitter

Autores: Adriano Gomes da Silva
Marçal Nunes de Oliveira
Venicius Bregalda

Introdução

Howard Phillips Lovecraft

Escritor Estadunidense (Gênero Horror)

- Considerado um dos *pais do gênero*
- Possui uma prolífica carreira literária



Introdução

Utilizando a rede social Twitter,

e as hashtags:

#HPLovecraft;

#HowardPhillipsLovecraft;

#Lovecraft,

pretende-se responder a seguinte hipótese:

Objetivo

Qual o conto do autor (entre 5), presente no livro 'The Best of H.P. Lovecraft Bloodcurdling Tales of Horror and The Macabre', é o mais mencionado em 50k ou mais tweets ?

Metodologia

Os 5 contos escolhidos através do método aleatório sem reposição, foram:

- Pickman's Model
- The Shadow Out of Time
- The Hunter of The Dark
- The Dunwich Horror
- The Call of Cthulhu

Metodologia

Por que esse livro?

Quando da realização da pesquisa, era o mais vendido no site Amazon.

Amazon Best Sellers
Our most popular products based on sales. Updated hourly.

[Any Department](#)

Best Sellers in 488907010 - Lovecraft, H. P.

- 

1. The Best of H.P. Lovecraft
H.P. Lovecraft
★★★★☆ 214
Paperback
\$12.80 ✓prime
- 

2. The Watchers Out of Time
H.P. Lovecraft
AND AUGUST DERLETH
★★★★☆ 36
Paperback
\$11.20 ✓prime
- 

3. The Horror in the Museum
H.P. Lovecraft
★★★★☆ 28
Paperback
\$16.93 ✓prime
- 

4. Nyarlathotep
- 

5. The Whisperer in the Walls
- 

6. At the Mountains of Madness

Metodologia

Linguagem:

Python 3.6.3

Bibliotecas:

- Crypto
- Pymongo

Armazenamento:

MongoDB 2.6.10

Ferramenta de Extração de Tweets elaborada como TCC do aluno
Édipo Deon Terra

Metodologia

Algoritmo para criptografia utilizado foi o AES 256, disponibilizado pela biblioteca crypto.

Três índices para a consulta foram criados (além do índice ID padrão que o próprio MongoDB cria) : índice pro tweet, índice para ano e índice para mês.

Metodologia

Algoritmo para criptografia utilizado foi o AES 256, disponibilizado pela biblioteca crypto.

Três índices para a consulta foram criados (além do índice ID padrão que o próprio MongoDB cria) : índice pro tweet, índice para ano e índice para mês.

- MongoDB permite a criação de índices *texto* para a consulta por texto no corpo do campo de um registro (*Full text search*)

Metodologia

```
def buscaTextual(self, textoBusca, conta=False): # busca textual nos TWEETS
    global posts
    if(conta==False):
        listaRetorno = []
        for entrada in posts.find({"$text": {"$search":"\""+textoBusca+"\""}}):
            anoDecrypt = str(self.cipher.decrypt(entrada["ano"]).decode("utf-8"))
            objeto = { 'ano': anoDecrypt, 'tweet':entrada["tweet"], 'mes':entrada["mes"]}
            listaRetorno.append(objeto);
        return listaRetorno
    else:
        return (posts.find({"$text": {"$search":"\""+textoBusca+"\""}})).count();|
```

```
buscaPickmanModel = ["Pickman's Model","Pickmans Model","pickman's Model","pickmans model"]
buscaShadowOut = ["The Shadow Out of Time","the shadow out of time","Shadow Out of Time","shadow out of time"]
buscaHaunter = ["The Haunter of the Dark","the haunter of the dark","Haunter of the Dark","haunter of the dark"]
buscaDunwich = ["The Dunwich Horror","Dunwich Horror","the dunwich horror","dunwich horror"]
buscaCthulhu = ["The Call of Cthulhu","Call of Cthulhu","the call of cthulhu","call of cthulhu"]
```

Resultados

Contos	<u>Quantidade de Menções</u>
Pickman's Model	144
The Shadow Out of Time	240
The Haunter of The Dark	148
The Dunwich Horror	520
The Call of Cthulhu	1796

fonte: autores

Resultados

Contos	<u>Quantidade de Menções</u>
Pickman's Model	144
The Shadow Out of Time	240
The Haunter of The Dark	148
The Dunwich Horror	520
<u>The Call of Cthulhu</u>	<u>1796</u>

fonte: autores

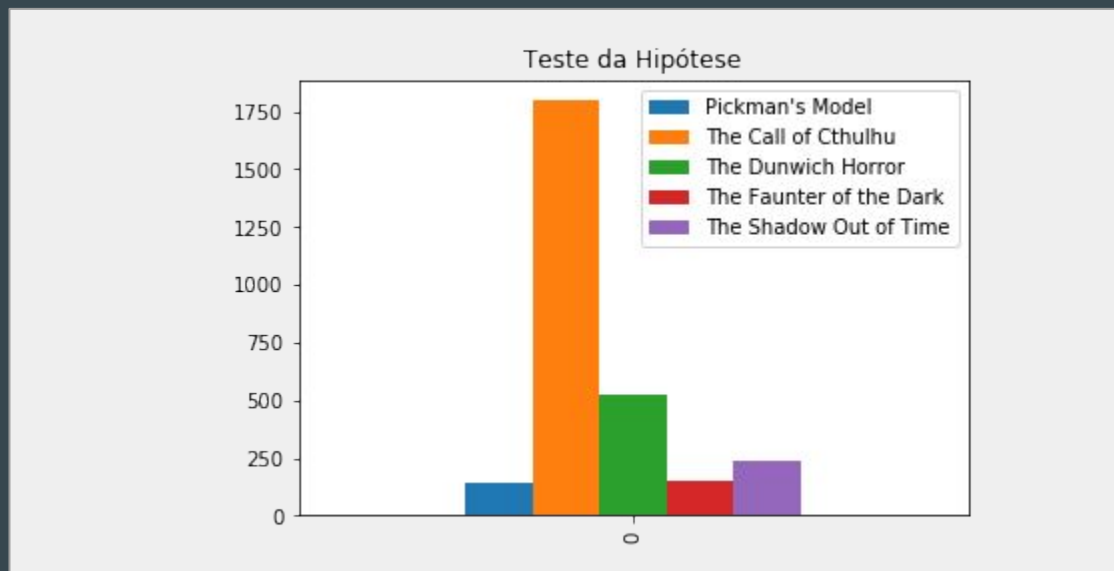
Resultados

Qual o conto do autor (entre 5), presente no livro 'The Best of H.P. Lovecraft Bloodcurdling Tales of Horror and The Macabre' é o mais mencionado em 50k ou mais tweets' ?

<u>The Call of Cthulhu</u>	<u>1796</u>
----------------------------	-------------

Resultados

Gráfico:



fonte: autores

Referências:

<http://www.sitelovecraft.com/hplovecraft.php#Biografia>

<http://www.hplovecraft.com/>

<http://www.hplovecraft.com/life/biograph.aspx>

<https://pypi.python.org/pypi/pycrypto>

<https://api.mongodb.com/python/current/>

<https://docs.mongodb.com/>

[https://repositorio.ucs.br/xmlui/bitstream/handle/11338/1124/TCC Edipo Deon Terra .pdf?sequence=1&isAllowed=y](https://repositorio.ucs.br/xmlui/bitstream/handle/11338/1124/TCC%20Edipo%20Deon%20Terra.pdf?sequence=1&isAllowed=y)