
Consulta - Sites de Jogos

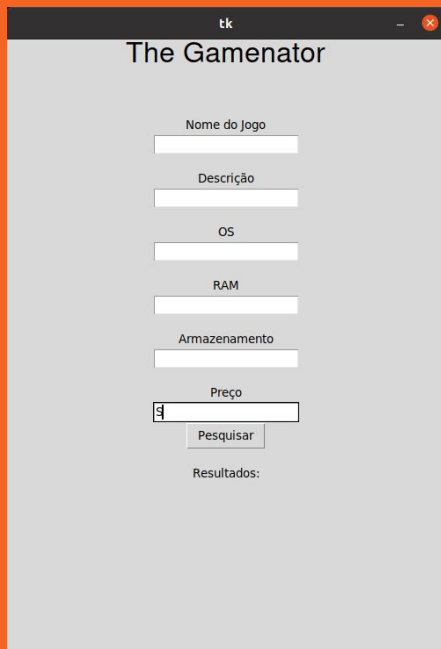
Italo Soares

2 scripts:

1. Lê os HTMLs e extrai informação para JSONs
2. Lê os JSONs e cria o arquivo de posting

Criação de posting com contagem manual de frequência e CountVectorize do SKlearn. Stemming foi usado para gerar índice invertido onde os termos estejam normalizados. Tamanho dos arquivos de índice invertido.

```
% python3 index.py  
Size of file Frequency: 25.90 kB  
Size of file TwoTerms: 35.34 kB  
Size of file Frequency Compressed: 25.48 kB  
Size of file TwoTerms Compressed: 34.92 kB
```



- Usando correlação de Spearman para comparar rankings.
- Uso de TFIDF opção, de acordo com uma flag no código.
- Interface feita com tkinter
- Query busca por strings sanitizadas sem acento, símbolos, capitalização, etc
- General Query busca por strings e atributos adicionais