Composição de Programas em Python

Tarefa 3.4 - Soup + MP

Data de entrega: domingo, 31 out 2021, 21:00
Arquivos requeridos: Driver.py, todo.py, exemplo/most-covered-tracks_1, exemplo/most-remixed-tracks_1, exemplo/most-sampled-tracks_1 (Baixar)
Tamanho máximo de arquivo carregado: 1 MiB
Tipo de trabalho: Trabalho individual

Neste exercício, assim como nos anteriores, você deverá completar algumas funções disponíveis em um arquivo `todo.py`. O arquivo é composto de uma série de funções que realizam diferentes fases de um sistema de map reduce paralelo. Seu trabalho vai ser processar uma grande quantidade de arquivos html usando Soup. Em particular, os arquivos inicialmente estão compactados via tar.gz. O código abaixo ensina como tem um esqueleto de uma função mapreduce.

Para testar em sua máquina baixe o arquivo: **dcc.ufmg.br/~flaviovdf/dados.tar.gz**

```
import multiprocessing as mp
import tarfile
tar = tarfile.open("filename.tar.gz", "r:gz")
# f = tar.extractfile(member) pega o arquivo. veja o getmembers abaixo
with mp.Pool(4) as pool:
    # extrai os arquivos e recebe um dicionário de {cantor: samples+remix+cover}
    resultado_intermed = pool.imap_unordered(extract_and_process,  tar.getmembers())
    final = reduce(merge_function, resultado_intermed)
```

Você deve implementar (em breve falo do arquivo):

- `extract_and_process`: vai extrair um arquivo do .tar.gz e retornar um dicionário de artist: popularidade. Retorne uma classe Counter
- `merge_function`: combina todos os sorts anteriores. Retorne uma classe Counter
- map_reduce: retorna o resultado final. Retorne uma classe Counter

Lembre-se do primeiro VPL deste módulo. É a mesma ideia, porém em paralelo.

Cada arquivo HTML contém uma lista de quantas vezes uma música foi Sampleada, Remixada ou Covered. Seu trabalho usando Soup é realizar a soma destes 3 valores por ARTISTA//BANDA. Por exemplo, ao processar o arquivo abaixo cada artista tem uma contagem de 14. Na pasta exemplos existem alguns HTML para vocês entenderem o formato.

- 91 You Like Me - Paris Jones You Like Me - Paris Jones Remixed 14 times
- 92 I Turn to You - Melanie C I Turn to You - Melanie C Remixed 14 times
- 93 Reality - Lost Frequencies feat. Janieck Devy Reality - Lost Frequencies feat. Janieck Devy Remixed 14 tim
- 94 Remember the Time - Michael Jackson Remember the Time - Michael Jackson Remixed 14 times
- 95 Finally - CeCe Peniston Finally - CeCe Peniston Remixed 14 times
- 96 Fade - Kristine W Fade - Kristine W Remixed 14 times
- 97 The Rising - Five Knives The Rising - Five Knives Remixed 14 times
- 98 Big Time Sensuality - Björk Big Time Sensuality - Björk Remixed 14 times
- 99 What Would We Do - DSK What Would We Do - DSK Remixed 14 times
- 100 Relax - Frankie Goes to Hollywood Relax - Frankie Goes to Hollywood Remixed 14 times