

TODO titulo

Trabalho elaborado no âmbito da Unidade Curricular de Armazenamento para Big Data do 2º ano da Licenciatura de Ciência de Dados do Instituto Universitário de Lisboa ISCTE

André Plancha; 105289

Another Author; Num

07/16/2020

Introdução

A Associação de Tenistas Profissionais (*ATP*) é um órgão de ténis profissional masculino, organizando torneios do desporto globalmente. A organização contém na sua base de dados um conjunto de jogos e jogadores que participaram em torneios pelo menos desde 1914, e incluem todos os grandes torneios do circuito masculino, incluindo os torneios de Grand Slam. O objetivo deste trabalho será limpar e preparar os dados de um modelo não-relacional para um modelo-relacional, para que possa ser utilizado em análises posteriores.

Importação dos dados

Para o nosso projeto foi-nos provisionado o ficheiro *atpplayers.json*, que contém os jogos feitos pelos jogadores. Para importar este ficheiro, foi utilizado o comando *mongoimport*:

```
mongoimport `
  --db atp `
  --collection games `
  --file "$pwd\data\atpplayers.json"
```

Isto significa que foi criada uma base de dados chamada *atp*, e uma coleção chamada *games*, que contém os dados do ficheiro *atpplayers.json*.

```
use atp;
db.games.find({}, {_id:0}).limit(5);
```

Born	Belgrade, Serbia	Belgrade, Serbia	Belgrade, Serbia	Belgrade, Serbia	Belgrade, Serbia
Date	2022.02.21 - 2022.02.26	2021.08.30 - 2021.09.12	2021.11.15 - 2021.11.21	2021.08.30 - 2021.09.12	2021.11.01 - 2021.11.07
GameRank	26	145	5	121	
GameRound	Round of 16	Round of 128	Round Robin	Round of 64	Round of 64
Ground	Hard	Hard	Hard	Hard	Hard

Hand	Right- Handed, Two-Handed Backhand	Right- Handed, Two-Handed Backhand	Right- Handed, Two-Handed Backhand	Right- Handed, Two-Handed Backhand	Right- Handed, Two-Handed Backhand
Height	188	188	188	188	188
Location	Dubai, U.A.E.	New York, NY, U.S.A.	Turin, Italy	New York, NY, U.S.A.	Paris, France
Oponent	Karen Khachanov	Holger Rune	Andrey Rublev	Tallon Griekspoor	bye
PlayerName	Novak Djokovic	Novak Djokovic	Novak Djokovic	Novak Djokovic	Novak Djokovic
Prize	\$2,794,840	\$27,200,000	\$7,250,000	\$27,200,000	5,207,405
Score	63 76	61 67, 62 61	63 62	62 63 62	null
Tournament	Dubai	US Open	Nitto ATP Finals	US Open	ATP Masters 1000 Paris
WL	W	W	W	W	

nota: Foi retirado a coluna LinkPlayer para melhor visualização

Info dump (TODO tratar e tirar antes de entregar)

count: 1308835 ## colunas * PlayerName: Nome do jogador * Born: onde o jogador nasceu (cidade, pais) * Height: altura do jogador (cm) * Hand: mão dominante, backhand * LinkPlayer: link para a página do jogador * Tournament : nome do torneio * Location : cidade, pais do torneio (atencao q US está [sitio, estado, U.S.A]) * Date: periodo de tempo do jogo (AAAA.MM.DD - AAAA.MM.DD) (pode-se tartar mas n vai ajudar nas perguntas) * GameRound: fase do jogo (pode-se tratar mas n vai ajudar nas perguntas) * GameRank: ATP Rankings do jogo? (int) * WL: vitoria ou derrota (W ou L) * Opponent: nome do oponente * Score: resultado do jogo (set scores); Este campo é mais complicado, mas pra as perguntas n é preciso desenvolver (ou entender) # Exemplo de uso do codigo (TODO tirar dps)

```
db.collection.find({}).pretty()
```

```
select * from table where x=2;
```