


Tokenização de Sentenças em Python

Extração de texto de páginas da web e
separação em sentenças

A large, dark blue, curved shape that starts from the bottom left and extends diagonally upwards towards the right, covering the bottom half of the slide.

Motivação

Por que tokenizar textos?

- Preparação para NLP (Processamento de Linguagem Natural)
- Facilita análise de dados textuais
- Usado em chatbots, análise de sentimentos, etc.

Objetivo do Projeto

Criar um programa que:

- Acesse um site e extraia seu conteúdo textual.
- Divida o texto em sentenças de forma automática.
- Exiba o texto original e o resultado da tokenização.

Tecnologias utilizadas

Linguagem: Python 3

Bibliotecas:

- requests → para fazer requisições HTTP
- BeautifulSoup (bs4) → para extrair o texto da página HTML
- re (expressões regulares) → para dividir o texto em sentenças

Texto original

E a tão requisitada Bola de Ouro da revista France Football foi mesmo para o francês Ousmane Dembélé, com toda a justiça. Realmente, foi a grande temporada de Dembélé, com conquistas coletivas incríveis que levaram ao seu desempenho individual extraordinário. É isso mesmo. Foi um prêmio individual, mas que chegou para um cara superimportante no mecanismo coletivo do melhor time do mundo, que é o PSG. O cara que também ganhou o prêmio como melhor treinador do mundo, o espanhol Luís Enrique, tem m

Sentenças tokenizadas

1. E a tão requisitada Bola de Ouro da revista France Football foi mesmo para o francês Ousmane Dembélé, com toda a justiça.

2. Realmente, foi a grande temporada de Dembélé, com conquistas coletivas incríveis que levaram ao seu desempenho individual extraordinário. É isso mesmo.

3. Foi um prêmio individual, mas que chegou para um cara superimportante no mecanismo coletivo do melhor time do mundo, que é o PSG.

4. O cara que também ganhou o prêmio como melhor treinador do mundo, o espanhol Luís Enrique, tem muito a ver com essa conquista do craque francês.

5. Anteriormente, Dembélé era um jogador totalmente individualista, que tentava fazer jogadas de efeito para si e não para sua equipe.

6. Luís Enrique chegou ao PSG e foi logo colocando a casa em ordem, não se opondo a quem queria sair, como Messi, mas dando linha para quem não se enquadraria no seu esquema coletivo de jogar.

7. Primeiro, foi Neymar que ele dispensou, e depois deu portas abertas também para Mbappé.

8. Juca Kfourti Todos sempre dizem que o Brasileiro está manchado Amanda Klein Governo cede e corre para salvar MP que taxa bets Raquel Landim No plano de Valdemar e Ciro, Caiado sobra Felipe Salto Precisamos celebrar e debater a Constituição Mas, com Ousmane Dembélé, quis conversar e tentar fazer com que ele entendesse o que tinha que fazer.

Sentenças tokenizadas

9. Não foi logo de cara que o vencedor da Bola de Ouro entendeu qual seria o seu papel dentro do mecanismo de grupo formado para que todos fossem importantes e não para correrem para um só jogador.

10. Ficou bravo, porém, com o tempo foi se enquadrando e chegou ao auge de ser determinante tanto coletivamente quanto individualmente.

11. Dembélé não é o mais importante do time, só do PSG, porque ninguém é; todos são, e esse é o segredo do sucesso do futebol atual.

12. Desequilibrou em diversas partidas da temporada, durante as conquistas que teve: Campeonato Francês, Copa da França, mas no principal, o inédito título foi soberbo, ajudando demais o PSG a ganhar sua primeira Champions League.

13. Eu achava que seria ele mesmo o vencedor, e para muitos, era o grande favorito por tudo que fez na temporada.

14. Mas parece que a próxima Bola de Ouro já tem um candidato fortíssimo, que é Lamine Yamal, que ganhou pela segunda vez consecutiva o prêmio Kopa como melhor jogador sub-21 do mundo e, de quebra, ficou em segundo lugar na Bola de Ouro.

15. Se eu votasse, meu voto seria para Ousmane Dembélé.

16. Texto em que o autor apresenta e defende suas ideias e opiniões, a partir da interpretação de fatos e dados. ** Este texto não reflete, necessariamente, a opinião do UOL.

17. O autor da mensagem, e não o UOL, é o responsável pelo comentário.

Sentenças tokenizadas

18. Leia as Regras de Uso do UOL.

19. Escolha de Trump para negociar com Brasil preocupa setor privado nacional Postos licenciados pelo Corinthians são alvos de megaoperação contra PCC Carros de luxo: com reação da Audi, veja os mais vendidos em setembro Bolsonaro não é citado em telefonema de 30 minutos entre Lula e Trump Vini Jr. e Virginia surgem juntos em vídeo de dancinha; assista Qual árbitro que prejudicou seu time que você tem mais bronca?

20. Os garimpeiros do VAR atacam novamente 'Não foi inexperiente': Casão defende Danilo em lance de expulsão 'Cara de campeão': Colunistas exaltam vitória do Palmeiras Flamengo sentiu a pressão contra o Bahia?

21. Colunistas debatem Filipe Luís errou em derrota do Flamengo?

22. Colunistas opinam Que Choque-Rei foi esse?

23. Erro da arbitragem não tira brilho do Palmeiras Luiz Gustavo, um cara que ama o que faz O dia que tive uma das maiores emoções da minha vida Flamengo x Cruzeiro: em 1993, vi um fenômeno surgir de dentro do campo Casão detona pênalti marcado contra o Corinthians: 'Garimpeiros do VAR' Com Flaco López e Vitor Roque em campo o espetáculo está garantido Assine UOL Telefone 4003-6118 Capitais 0800 703 300 Demais localidades Baixe nossos apps 1996 - 2025 UOL - O melhor conteúdo.

24. Todos os direitos reservados.

25. Segurança e privacidade

Obrigado pela atenção!

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.