



Processamento de linguagem natural com Python

Coordenador
Fernando Ferreira





Institucional

Reitor

Prof. Eduardo Ramos

Eduardo.ramos@lnfnet.edu.br

Coordenação do Curso

Prof. Fernando Ferreira

Fernando.gFerreira@lnfnet.edu.br

Gerente Acadêmica

Ana Curi

ana.curii@lnfnet.edu.br

Coordenador/Professor



Prof. Fernando Ferreira

fernando.gferreira@infnet.edu.br

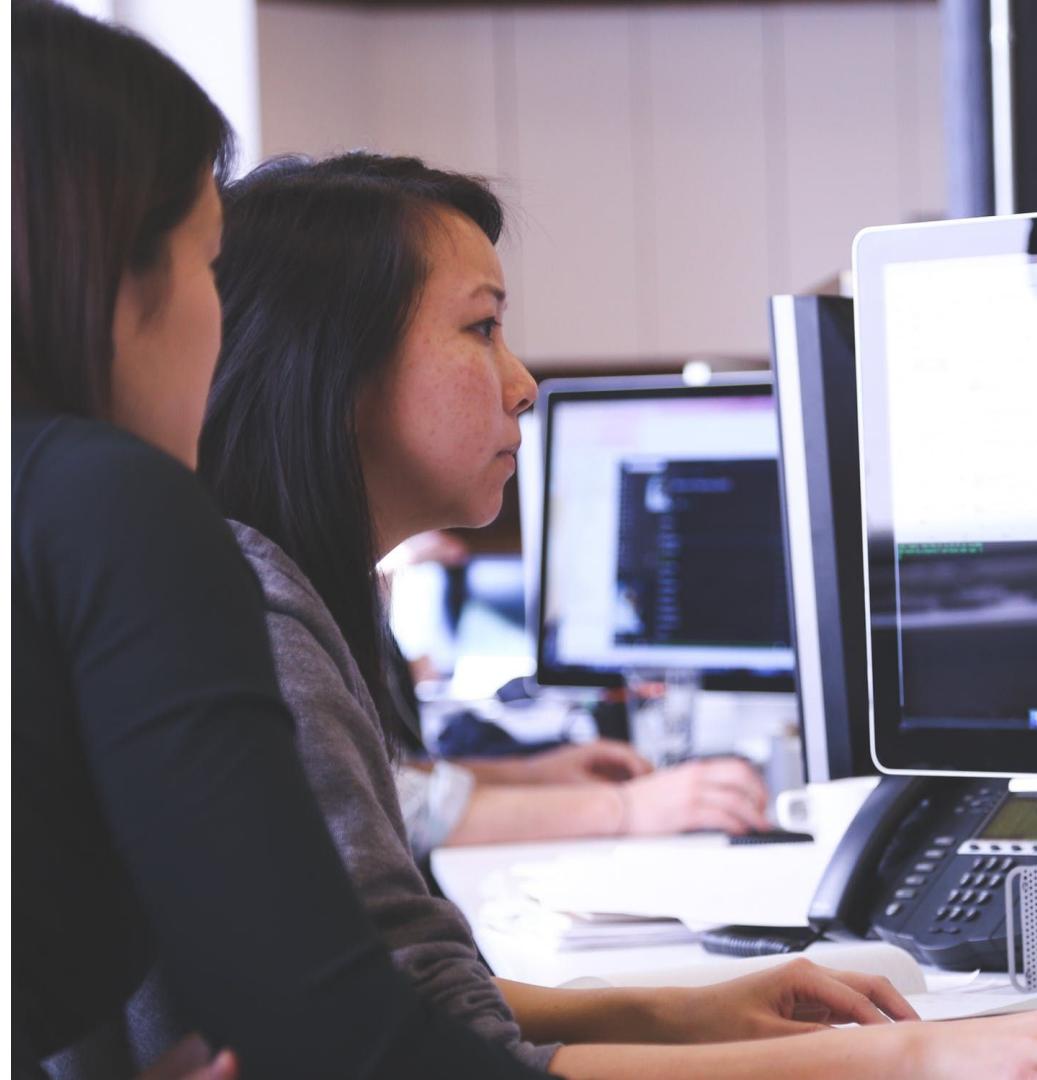
<https://www.linkedin.com/in/nandoferreira/>

Mini-bio

- Doutor pelo Programa de Engenharia Elétrica da Universidade Federal do Rio de Janeiro (Coppe-UFRJ) - ênfase em Inteligência Computacional.
- Sócio-fundador da empresa Twist (<https://twist.systems/>) - residente do Parque Tecnológico da UFRJ.
- Coordenador do MBA de Data Science e do MIT de Inteligência Artificial do Instituto Infnet.
- Professor de disciplinas de desenvolvimento ágil de software na graduação da Infnet. Professor do MBA de Business Analytics e Big Data da Fundação Getúlio Vargas.
- Membro da colaboração internacional CERN/UFRJ.
- [Twist] Empresa membro-nato do CDIA.Rio - Rede de Ciência de Dados & Inteligência Artificial do Rio de Janeiro

Mineração de texto e processamento de linguagem natural

Introdução



População
Mundial



7.5 Bilhões
urbanização:
54 %

Usuários
Internet



3.8 Bilhões
penetração:
50 %

Presença Redes
Sociais



2.8 Bilhões
penetração:
37 %

Aparelhos
Móveis



4.9 Bilhões
penetração:
66 %

Usuários de Redes
Sociais Móveis



2.5 Bilhões
penetração:
34 %

Em 2017, **139.1 milhões de pessoas** tinham acesso a internet (66% da população)



Brasileiros passam por dia, **9 horas e 14 minutos** surfando na internet.

127 milhões de usuários ativos no Whatsapp

Redes sociais Brasil

220 milhões de smartphones ativos

92% dos usuários do Facebook acessam a rede social via mobile

3 hours são gastas online acessando redes sociais

130 milhões de brasileiros (62% da população) estão conectados através das redes sociais

Open Data

DATA.GOV

DATA TOPICS IMPACT APPLICATIONS DEVELOPERS CONTACT

The home of the U.S. Government's open data

Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more.

BRASIL Serviços Participe Acesso à informação Legislação Canais

Ministério da Transparéncia e Controle Geral da União

Portal da Transparência

GET STARTED SEARCH OVER 190,000 DATA SETS

Credit Card Complaints

BROWSE TOPICS

Agriculture Climate Consumer Ecosys

Health Local Government Manufacturing Marit

CONSULTAS Despesas Recetas Convênios Empresas Sancionadas Empresas Punitidas Entidades Impedidas Servidores Imóveis Funcionais Benefícios L10.599/02

DESPESAS RECEITAS CONVÊNIOS

Despesas - Empenho, liquidação e pagamento Abatimento diário (a partir de maio de 2010) Despesas com diárias e despesas com diárias financeiras do governo federal, com possibilidade de uso de despesas (empreito, reembolso ou pagamento)

Despesas - Pagamento

Transferências de Recursos

Consultas Temáticas

DADOS DO PORTAL - DESPESAS

1 Recursos registrados (mesmo): R\$ 20.588,11

1 Informações registradas: 2.449.179.928

sign in become a supporter subscribe search

jobs dating more International edition

the guardian

home Open journalism Datablog

Open data journalism

Why does data journalism need to be open journalism too? Simon Rogers explains how we do it on the Datablog

More data journalism and data visualisations from the Guardian

Thursday 20 September 2012 08.00 BST

13 2

Simon Rogers

Can charities and non-governmental organisations use their information more effectively? Photograph: Deco / Alamy/Alamy

Pare de seguir leads que não produzem resultados

Permita que o Pipedrive lhe mostre o caminho

Teste gratuitamente

Data is everywhere: from governments publishing billions of bytes of the stuff, to visual artists creating new concepts of the world through to companies building businesses on the back of it.

Processamento de Linguagem Natural



NLP são **métodos** para computadores
analisarem, entenderem e obterem
significados para a **linguagem humana** de uma
maneira inteligente e útil

Quais são as tendências atuais?

Existe crise?





Jornalismo de Dados Como apoiar as informações em Dados?



Como achar influenciadores? Padrões de Usuários?

Mensurar Sentimentos



Geolocalizar



Dados não-estruturados, enriquecidos

The screenshot shows a news article from DW (Deutsche Welle) on their website. The headline reads: "Google buys stake in SpaceX to beam Internet from outer space". The text below states: "Google has secured a 10 percent stake in the private space exploration company SpaceX along with another investor. Reports suggest the money is aimed at hooking up far-off areas with Internet access from outer space." At the bottom, it says: "The most remote, barren corners of the globe may soon have Internet access." To the right of the article are two detailed infographics comparing SpaceX and Google.

SpaceX
Space Exploration Technologies Corp logo

Type	Private
Industry	Aerospace
Founded	2002
Founder	Elon Musk
Headquarters	Hawthorne, California, USA
Key people	Elon Musk (CEO and Chief Designer) Gwynne Shotwell (President and COO) Tom Mueller (VP of Propulsion)
Services	Orbital rocket launch
Number of employees	3,000+ (Jan 2015) <small>(citation needed)</small>
Website	SpaceX.com

Google Inc.

Type	Public
Traded as	Class A: NASDAQ: GOOGL Class B supervoting; unlisted Class C nonvoting: NASDAQ: GOOG
Industry	Internet Computer software Telecoms equipment
Founded	September 4, 1995; 16 years ago Menlo Park, California ^{[1][2]}
Founder	Larry Page, Sergey Brin
Headquarters	Googleplex, Mountain View, California, U.S. ^[3]
Area served	Worldwide
Key people	Larry Page (CEO) Eric Schmidt (Chairman) Sergey Brin (Director of Google X and Special Projects) ^[4]

[Footnotes / references](#)

[Send us your feedback.](#)

[Print](#) Print this page

Permalink <http://dw.de/p/1ENVv>