TÉCNICAS PROCESAMIENTO DE LENGUAJE NATURAL

Presenta

RECONOCIMIENTO DE ENTIDADES NOMBRADAS (NAMED ENTITY RECOGNITION)

Por

Michelle Díaz

- **Michelle Díaz**
- Ingeniería en Informática
- +1 año aprendiendo sobre AI y NLP

michellediazvi

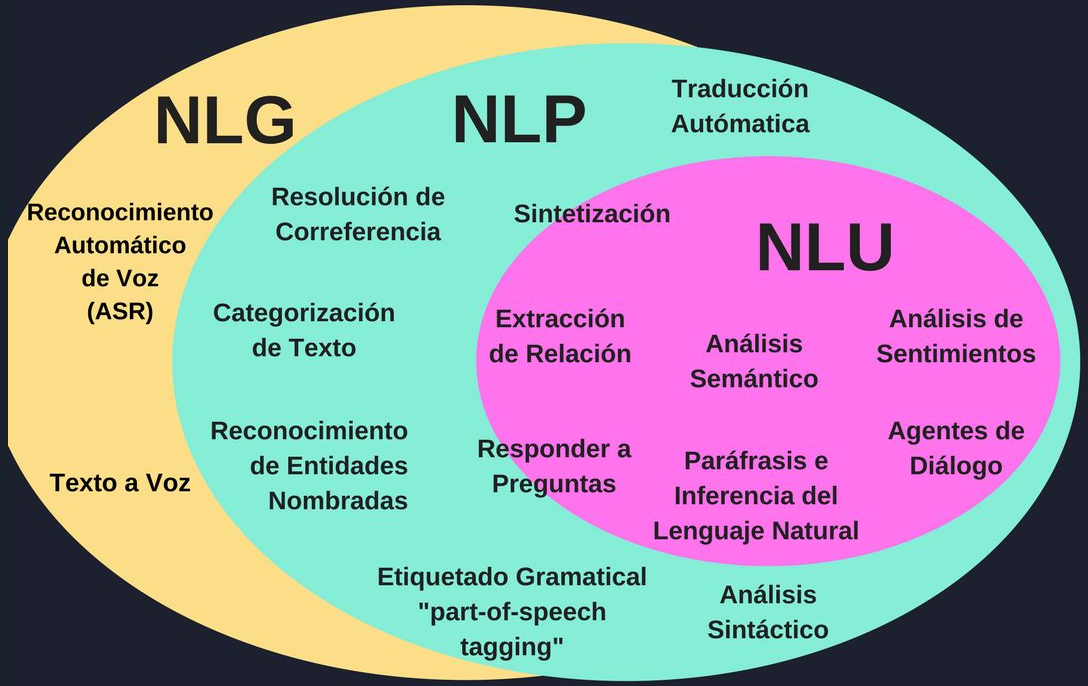
@MichDiaz_

@MichDiaz_

¿Qué es Procesamiento de Lenguaje Natural?

NLP v.s. NLU v.s. NLG

NLP - Procesamiento de Lenguaje Natural
NLU - Entendimiento de Lenguaje Natural
NLG - Generación de Lenguaje Natural



Fuente: https://nlp.stanford.edu/~wcmac/papers/20140716-UNLU.pdf

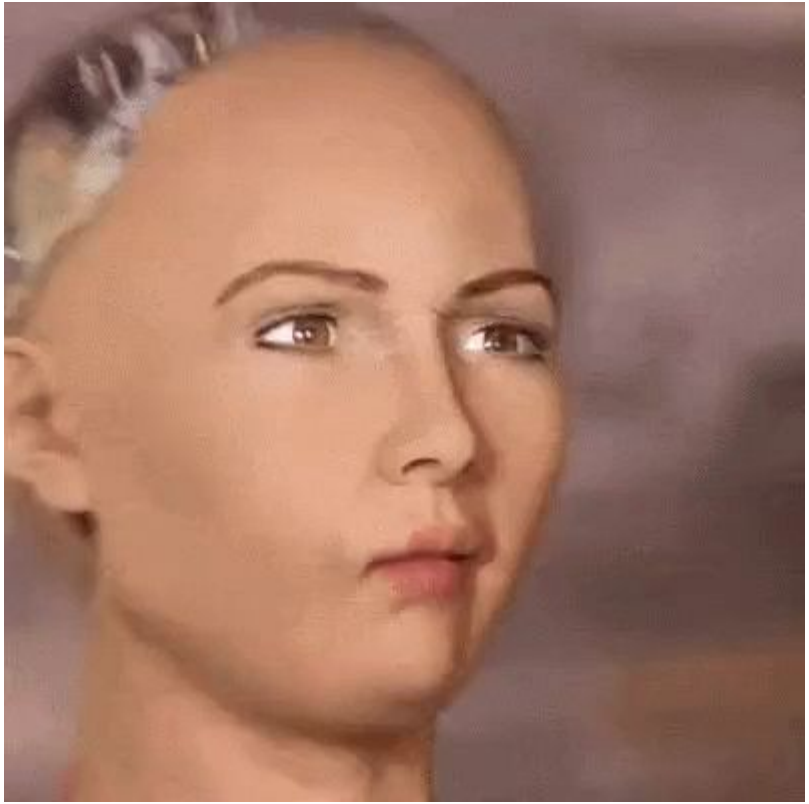
NLP v.s. NLU v.s. NLG

NLP - Procesamiento de Lenguaje Natural
NLU - Entendimiento de Lenguaje Natural
NLG - Generación de Lenguaje Natural

Fuente: https://nlp.stanford.edu/~wcmac/papers/20140716-UNLU.pdf

¿Qué aprenderemos?

- ¿Qué es NER?
- ¿Por qué usar NER?
- ¿Cómo desarrollar NER?
  - Datos
  - Evaluación
  - Herramientas
- API’s
- Casos de uso
- Tips
- Preguntas Frecuentes



¿Qué es NER?

STEPHEN HAWKING

Nació el 8 de enero de 1942 en Oxford, lugar al que expresamente se desplazaron sus padres, Isobel Hawking y Frank Hawking, investigador biológico, buscando una mayor seguridad para la gestación de su primer hijo, ya que Londres estaba siendo atacada por la Luftwaffe. Tiene además dos hermanas menores, Philippa y Mary, y un hermano adoptado, Edward.

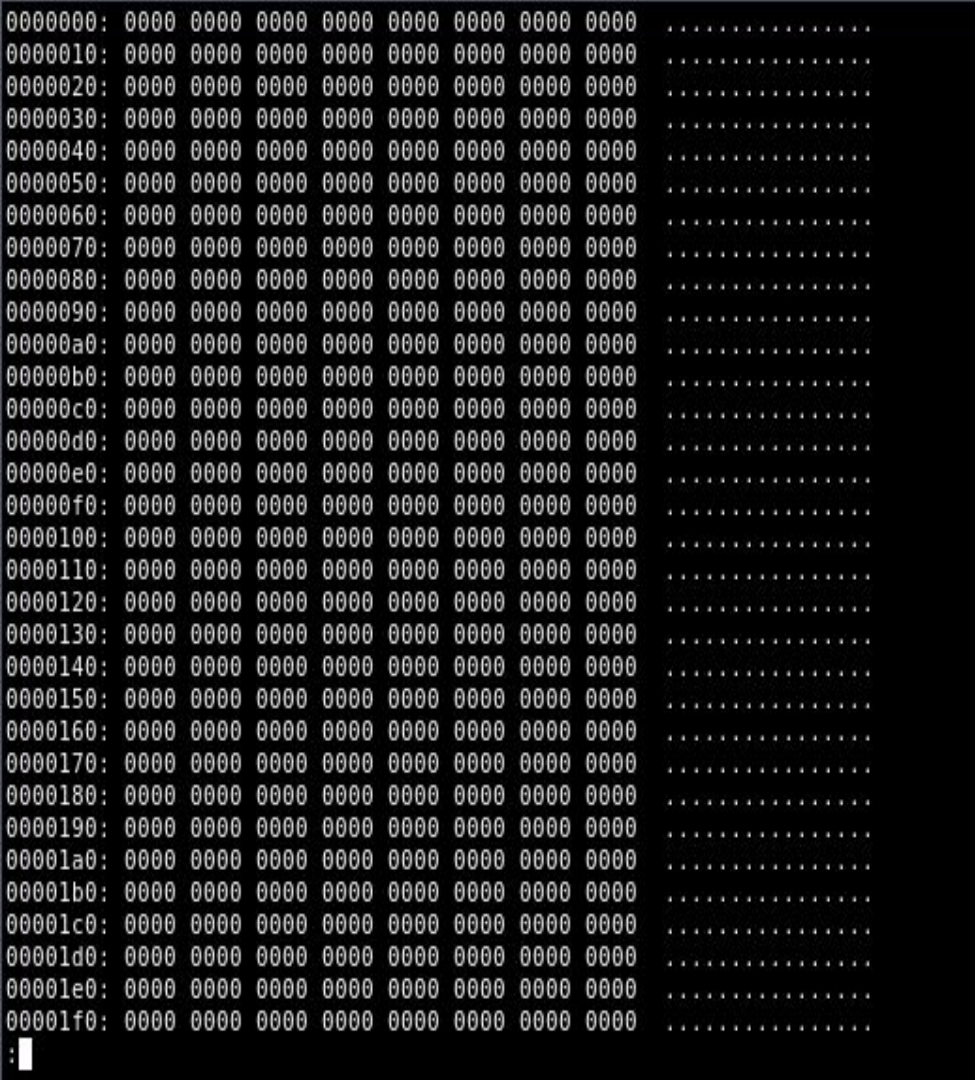
Después del nacimiento de Stephen, la familia volvió a Londres, donde su padre encabezaba la división de parasitología del National Institute for Medical Research. En 1950 se mudaron a St Albans, donde acudió al Instituto para chicas de St Albans (que admitía chicos hasta la edad de 10 años) y a los 11 años cambió al colegio homónimo, donde fue un buen estudiante aunque no brillante.

Fecha

Lugar

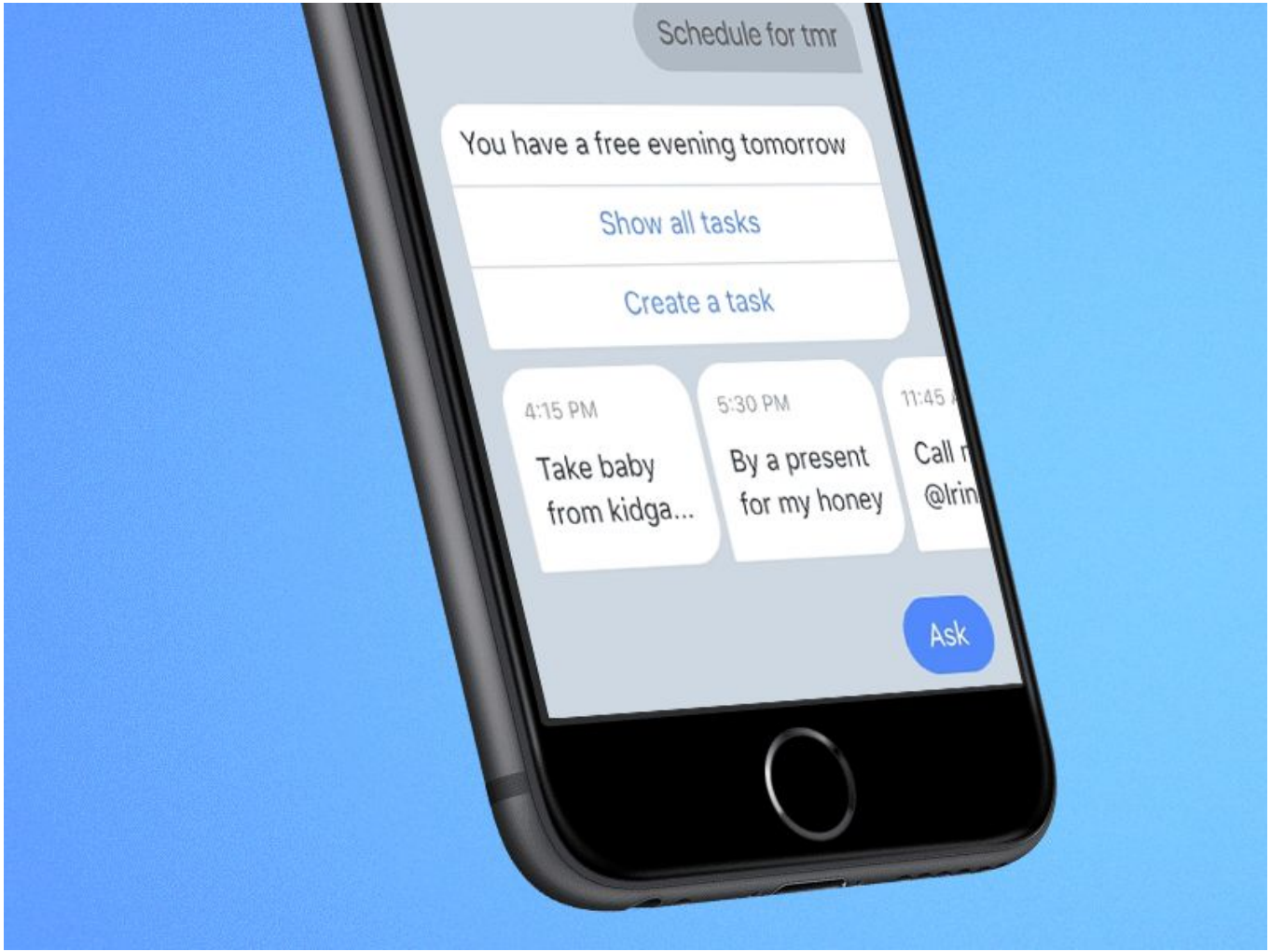
Persona

Organización



¿Por qué usar NER?

Para desarrollar y crear aplicaciones de alto nivel como:

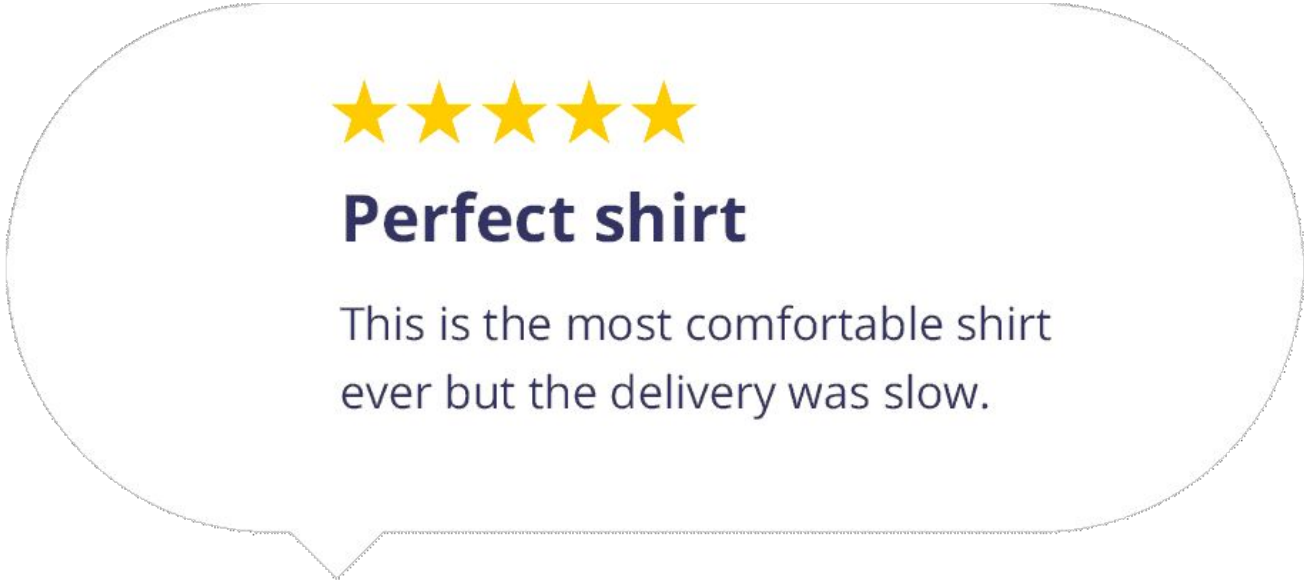

Schedule for tmr
You have a free evening tomorrow
Show all tasks
Create a task
4:15 PM
Take baby from kidga...
5:30 PM
By a present for my honey
Ask

CHATBOTS

QUESTION ANSWERING

SISTEMAS DE RECOMENDACIÓN

ANÁLISIS DE SENTIMIENTOS

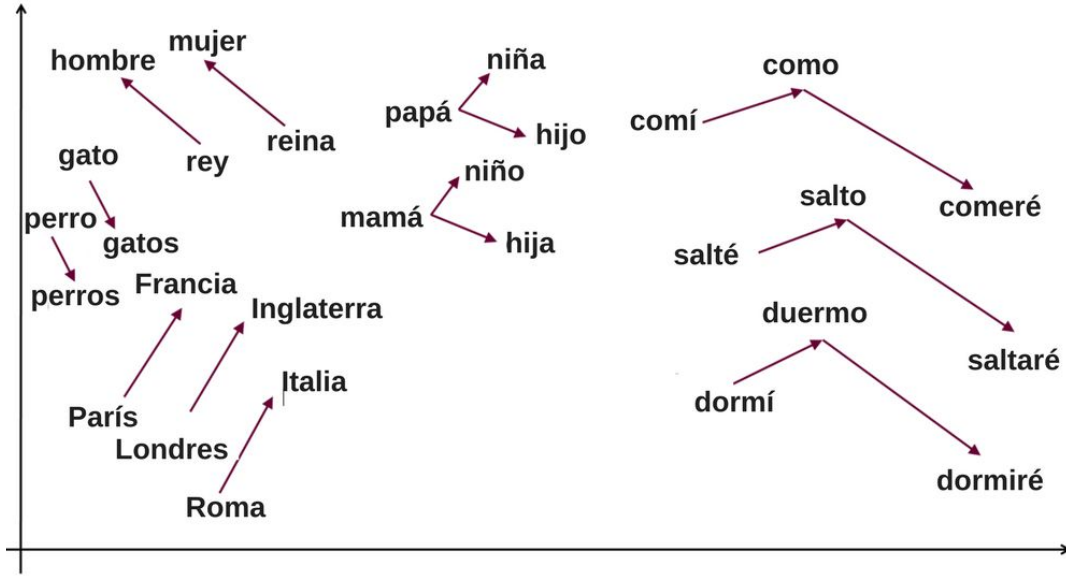


¿Cómo desarrollar NER? (Datos)

- **Representación de las palabras;** convertir palabras a “algo” que las computadoras puedan entender.

Sí… Pero, ¿Cómo?

- **Vectores de palabras**
  - Word2Vec
  - Glove



- **One-hot encoding**
  Por ejemplo, gato se puede representar como [.1233, .656575, .86864]

¿Cómo desarrollar NER? (Evaluación)

Precisión; **% de elementos seleccionados que son correctos**
Recall; **% de elementos correctos que se seleccionan**

Actual	Organización	Persona	Fecha	Persona	Organización
Predicción	Organización	Fecha	Fecha	Persona	Organización

Precisión = 80%
Recall = 80%

¿Cómo desarrollar NER? (Herramientas)

GEMSIM NLTK

In [2]:
```
# These are css/html style for good looking ipython notebooks
from IPython.core.display import HTML
css = open('c:/ml/style-notebook.css').read()
HTML('<style>{}</style>'.format(css))
```

Out[2]:

In [1]:
```
# -*- coding: utf-8 -*-
import gensim
import logging
import os
import nltk.data
import string
%matplotlib inline

logging.basicConfig(format='%(asctime)s : %(levelname)s : %(message)s', level=logging.INFO)

print ("PACKAGES LOADED")
```

```
C:\Anaconda2\envs\tensorflow-gpu\lib\site-packages\gensim\utils.py:855: UserWarning: detected Windows; aliasing chunk
ize to chunkize_serial
  warnings.warn("detected Windows; aliasing chunkize to chunkize_serial")
```

¿Cómo desarrollar NER? (Herramientas)

GloVe: Global Vectors for Word Representation

Jeffrey Pennington, Richard Socher, Christopher D. Manning

Introduction

GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.

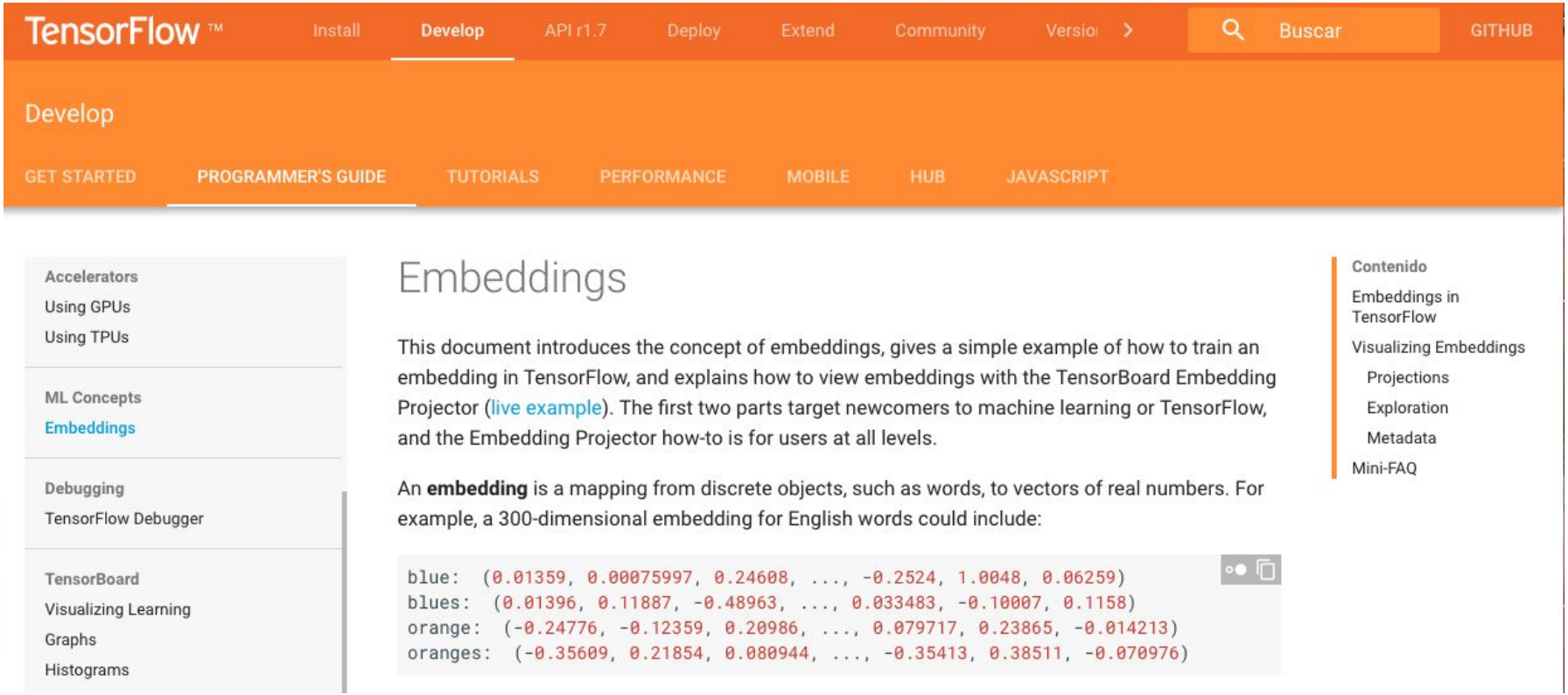
Getting started (Code download)

- Download the code (licensed under the Apache License, Version 2.0)
- Unpack the files: unzip GloVe-1.2.zip
- Compile the source: cd GloVe-1.2 && make
- Run the demo script: ./demo.sh
- Consult the included README for further usage details, or ask a question
- The code is also available on GitHub

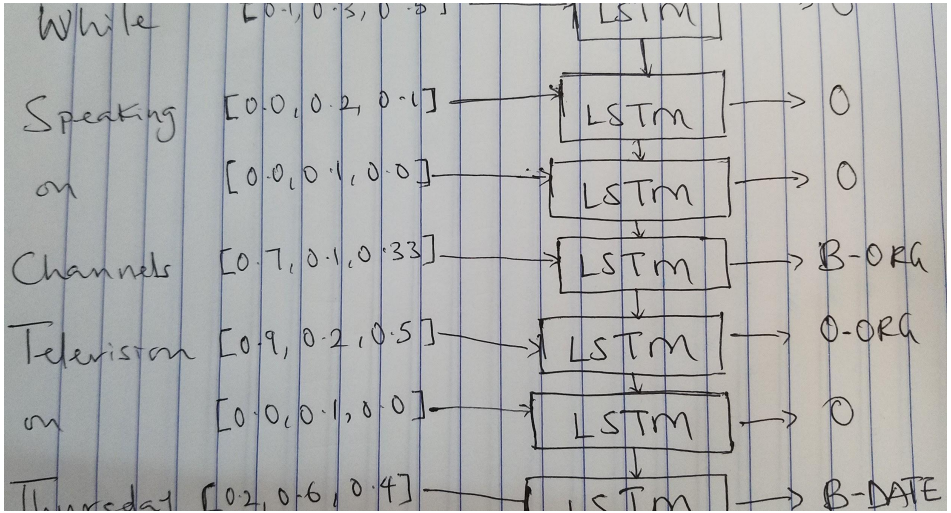
Download pre-trained word vectors

- Pre-trained word vectors. This data is made available under the Public Domain Dedication and License v1.0 whose full text can be found at: http://www.opendatacommons.org/licenses/pddl/1.0/.
  - Wikipedia 2014 + Gigaword 5 (6B tokens, 400K vocab, uncased, 50d, 100d, 200d, & 300d vectors, 822 MB download): glove.6B.zip
  - Common Crawl (42B tokens, 1.9M vocab, uncased, 300d vectors, 1.75 GB download): glove.42B.300d.zip
  - Common Crawl (840B tokens, 2.2M vocab, cased, 300d vectors, 2.03 GB download): glove.840B.300d.zip
  - Twitter (2B tweets, 27B tokens, 1.2M vocab, uncased, 25d, 50d, 100d, & 200d vectors, 1.42 GB download): glove.twitter.27B.zip
- Ruby script for preprocessing Twitter data

¿Cómo desarrollar NER? (Herramientas)

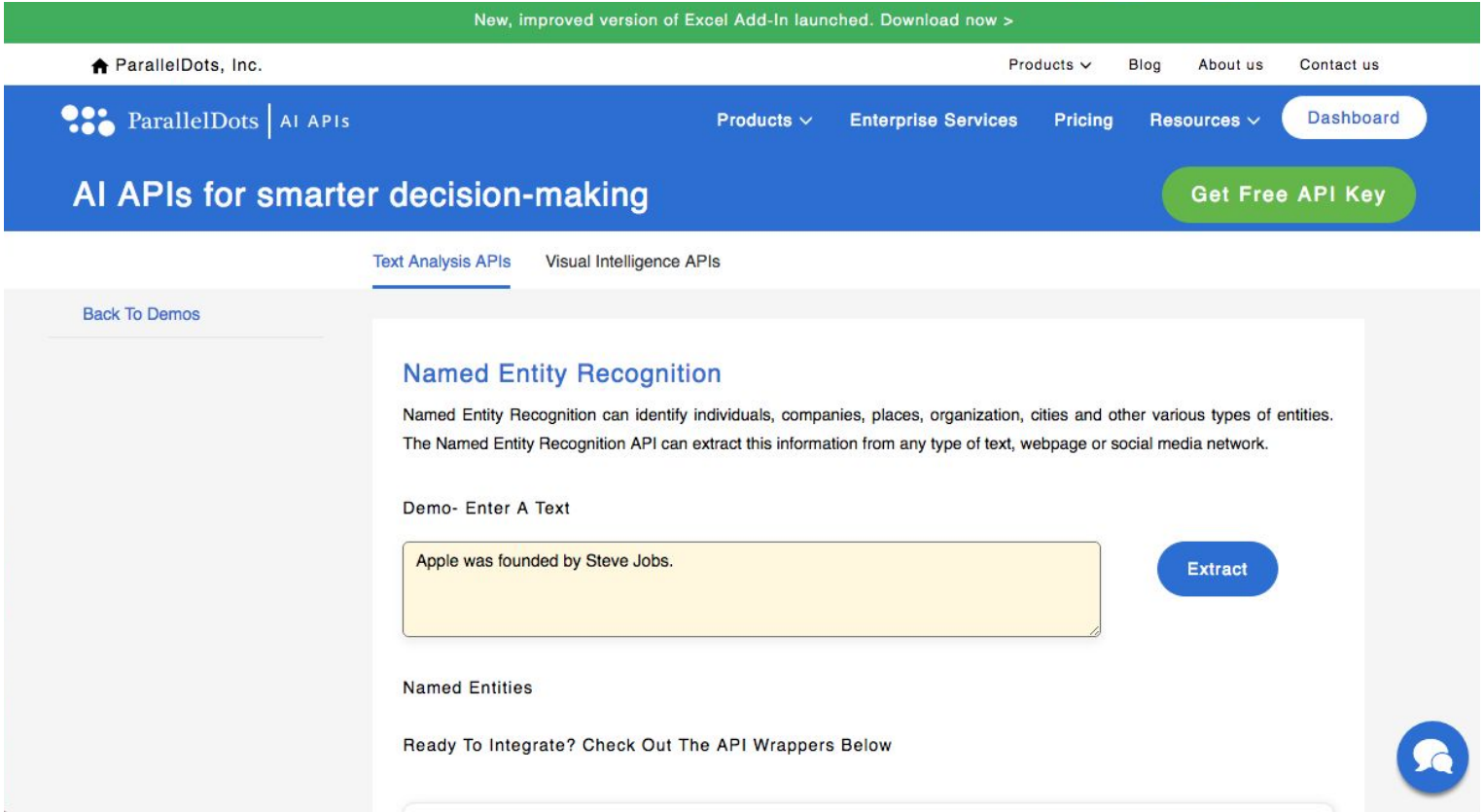


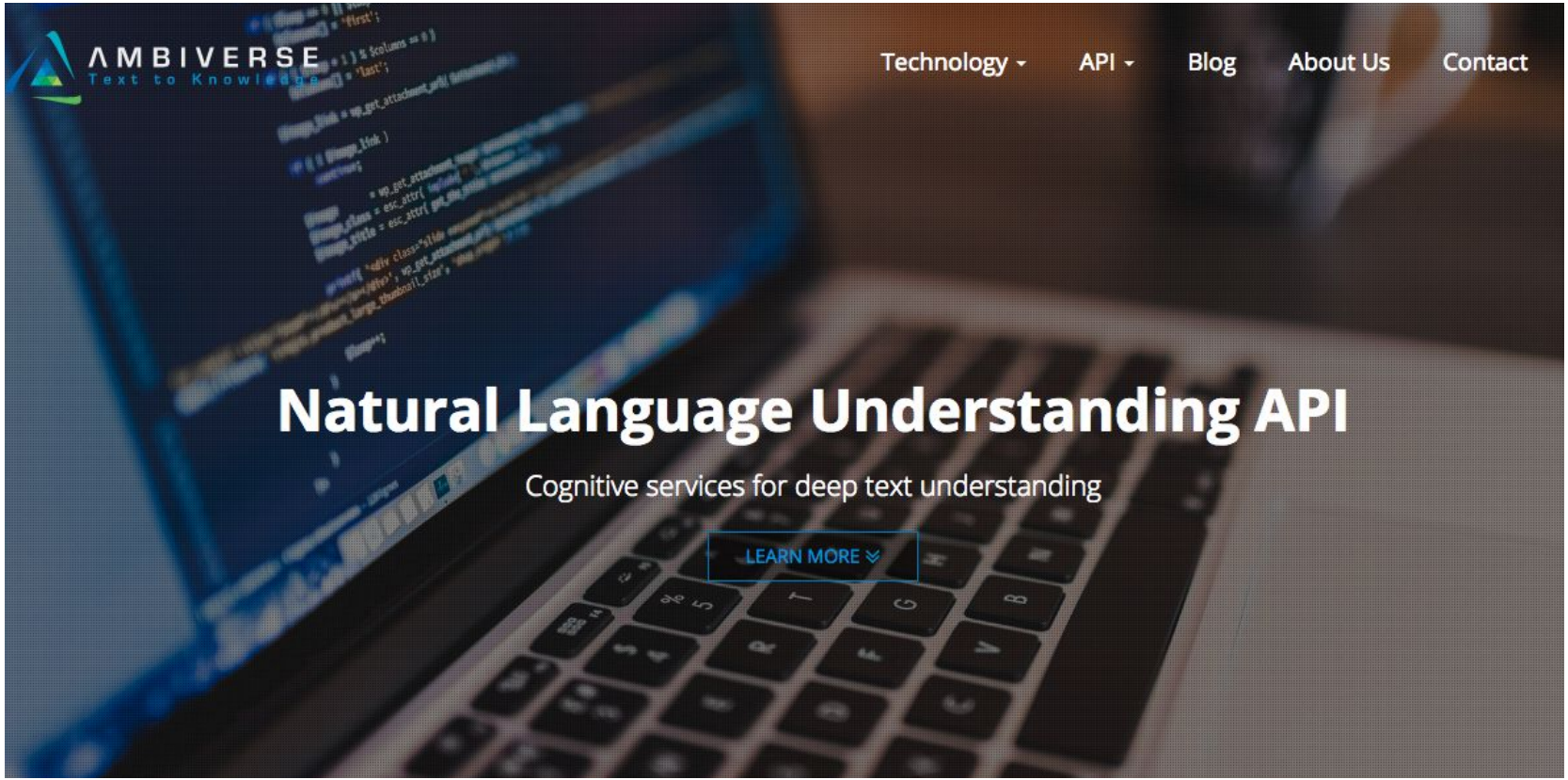
¿Cómo desarrollar NER? (Arquitectura)



- LSTM
- Word embeddings como input
- Etiquetas IOB como output

API’s


New, improved version of Excel Add-In launched. Download now >
ParallelDots, Inc.
Products
Blog
About us
Contact us
ParallelDots | AI APIs
Products
Enterprise Services
Pricing
Resources
Dashboard
AI APIs for smarter decision-making
Get Free API Key
Text Analysis APIs
Visual Intelligence APIs
Back To Demos
Named Entity Recognition
Named Entity Recognition can identify individuals, companies, places, organization, cities and other various types of entities. The Named Entity Recognition API can extract this information from any type of text, webpage or social media network.
Demo- Enter A Text
Apple was founded by Steve Jobs.
Extract
Named Entities
Ready To Integrate? Check Out The API Wrappers Below


AMBIVERSE
Text to Knowledge
Technology
API
Blog
About Us
Contact
Natural Language Understanding API
Cognitive services for deep text understanding
LEARN MORE

Google Cloud

Buscar CONSOLA

Por qué elegir Google | Productos | Soluciones | Launcher | Precios | Clientes | D >

PRUÉBALO GRATIS | CONTACTAR CON VENTAS

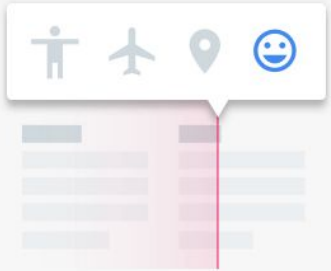
API NATURAL LANGUAGE DE CLOUD

Extrae información valiosa de textos no estructurados con el aprendizaje automático de Google

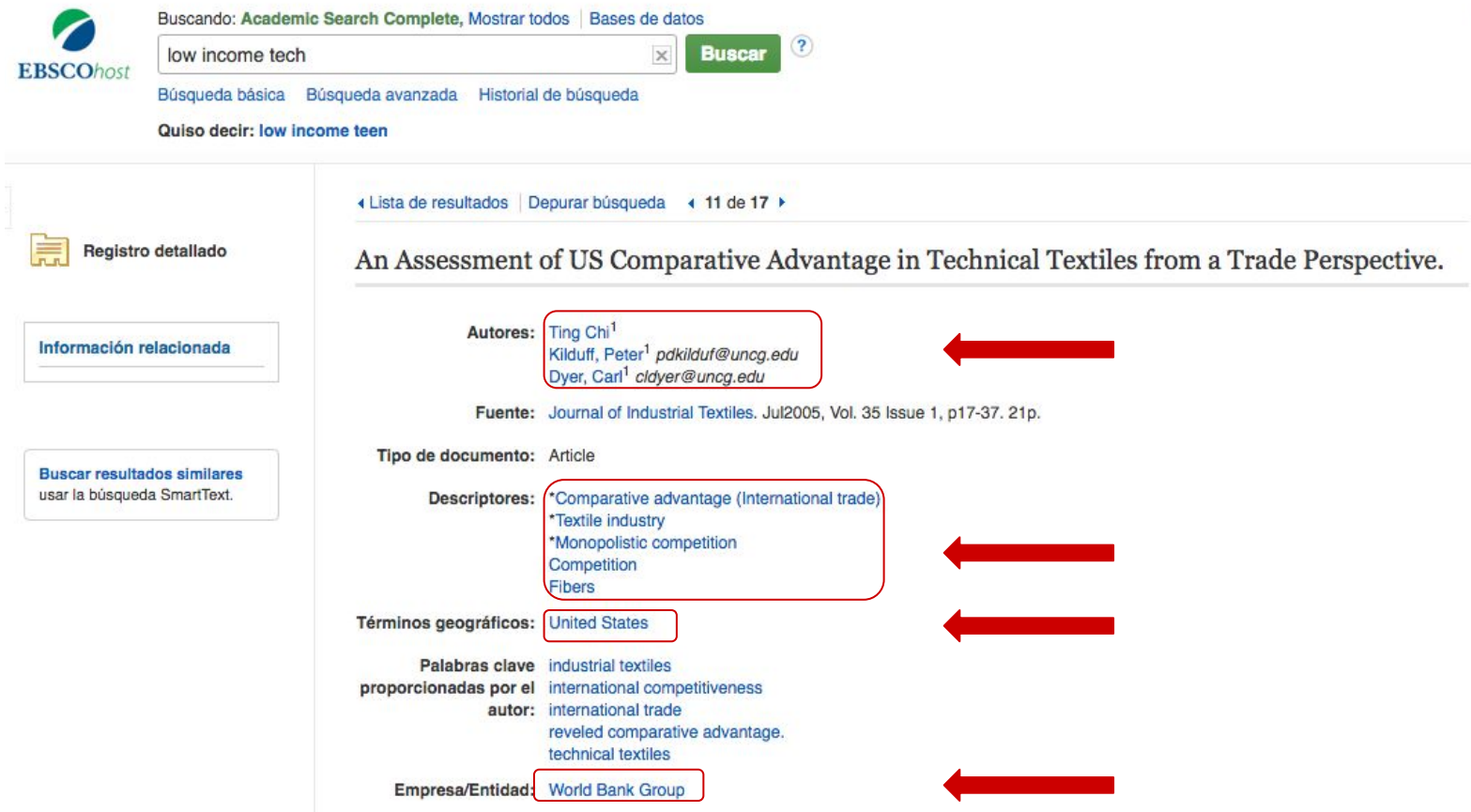
PRUÉBALO GRATIS | VER CONSOLA

Potente análisis de texto

La API Natural Language de Cloud descubre la estructura y el significado del texto mediante modelos de aprendizaje automático en una API REST fácil de usar. Puedes utilizarla para **extraer información** sobre personas, lugares, eventos y muchos elementos más que se mencionen en documentos de texto, artículos de noticias o entradas de blogs. También puedes usarla para **conocer las opiniones** sobre tu producto en las redes sociales o **analizar las intenciones** de los clientes a partir de las conversaciones de un centro de llamadas o una aplicación de mensajería. Es posible **analizar el texto que se**



Casos de Uso


EBSCOhost
Buscando: Academic Search Complete, Mostrar todos | Bases de datos
low income tech
Buscar
Búsqueda básica Búsqueda avanzada Historial de búsqueda
Quiso decir: low income teen
Registro detallado
Información relacionada
Buscar resultados similares
usar la búsqueda SmartText.
Lista de resultados | Depurar búsqueda 11 de 17
An Assessment of US Comparative Advantage in Technical Textiles from a Trade Perspective.
Autores: Ting Chi[1]
Kilduff, Peter[1] pdkilduf@uncg.edu
Dyer, Carl[1] cldyer@uncg.edu
Fuente: Journal of Industrial Textiles. Jul2005, Vol. 35 Issue 1, p17-37. 21p.
Tipo de documento: Article
Descriptores: *Comparative advantage (International trade)
*Textile industry
*Monopolistic competition
Competition
Fibers
Términos geográficos: United States
Palabras clave proporcionadas por el autor: industrial textiles
international competitiveness
international trade
reveled comparative advantage.
technical textiles
Empresa/Entidad: World Bank Group

Intelligent Machines

The GANfather: The man who's given machines the gift of imagination

By pitting neural networks against one another, Ian Goodfellow has created a powerful AI tool. Now he, and the rest of us, must face the consequences.

by Martin Giles February 21, 2018

One night in 2014, Ian Goodfellow went drinking to celebrate with a fellow doctoral student who had just graduated. At Les 3 Brasseurs (The Three Brewers), a favorite Montreal watering hole, some friends asked for his help with a thorny project they were working on: a computer that could create photos by itself.

Share

Tagged Ian Goodfellow, 10 Breakthrough Technologies 2018, neural networks, GAN, generative adversarial network

Algoritmos de búsqueda eficientes

Forbes Staff
abril 6, 2018 @ 4:37 pm

5 cursos gratuitos para aprender inteligencia artificial

La inteligencia artificial se encuentra presente en industrias como la financiera, telecomunicaciones, retail y farmacéutica.

También te puede interesar

Actualidad

El niño de 14 años que corrigió a IBM Watson y hoy es su 'asesor'

A su corta edad, Tanmay Bakshi ya es experto en Inteligencia Artificial y una especie de consej
una de las marcas...

Capital Humano

Confidencias | Tesla le muestra a Trump como debe ser el TLCAN

Capital Humano

Apple quiere recuperar el salón de clases, ¿aún está a tiempo?

Capital Humano

El camino hacia a las operaciones inteligentes

Capital Humano

La IA, el cambi disruptivo en la

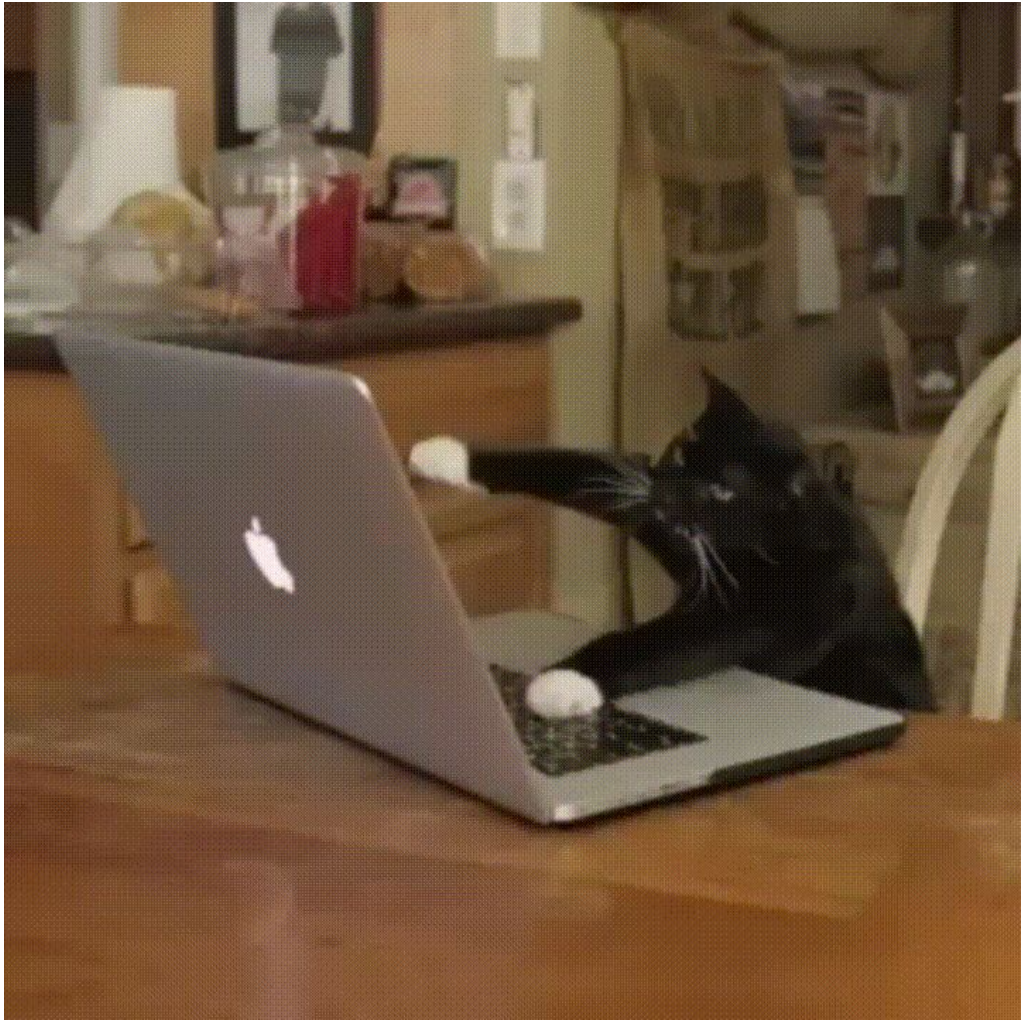
Atención al Cliente

Sandhya Advani
@sandyaadvani

@cromaretail please train your staff in croma bandra to provide correct details of customer support for Fitbit. The number given doesnt work

3:40 AM - Apr 16, 2017

1 See Sandhya Advani's other Tweets



TIPS

Estrategias para la extracción de datos:

- Estrategias basadas en la popularidad
- Estrategias lingüísticas
- Estrategias estadísticas
- Estrategias semánticas

El proceso de selección de la entidad está determinado por:

- Contexto
- Ambigüedad de los datos de origen / mapeo
- Precisión / fiabilidad de los datos de origen

Determinar todos los candidatos de mapeo de Entidades posibles:

- Análisis lingüístico (etiquetado POS)
- Normalización
- Codificación y ortografía
- Caracteres especiales (dependientes del idioma)
- Abreviaciones, acrónimos
- Ortografía dependiente del tope
- Nombres alternativos y sinónimos
- Mapeo de oraciones difusas

Estudiar las propiedades lingüísticas de un lenguaje es importante.

Preguntas Frecuentes

¡Gracias!