

Introducción General al PLN


Luis Espinosa, PhD

LingWars Madrid - 5 Octubre 2017

Índice

- Presentación
- ¿Qué es el PLN?
- Problemas fundamentales en PLN
- Aplicaciones
- Q&A

Presentación

- Un poco sobre mí
 - De qué sé más, y de qué menos.
- ¿Por qué nos interesa el PLN? Profesional, intelectual, diversión () , etc.
- Cuáles son los requisitos para hacer PLN?
 - Programación
 - Lingüística
 - Mates
 - Creatividad

¿Qué es el Procesamiento del Lenguaje Natural?

- El PLN es una rama de la Inteligencia Artificial orientada a “humanizar” la comunicación entre personas y máquinas.
- Está presente en prácticamente todos los productos digitales que usan lenguaje.
 - Buscadores, traductores, GPSs, sugerencias en editores, correctores, sistemas de *question answering*, minería de opiniones... y sus aplicaciones industriales son...

¿Qué es el Procesamiento del Lenguaje Natural?

- El PLN es una rama de la lingüística que se ocupa de la comunicación entre personas y máquinas.
- Está presente en prácticas cotidianas que usan lenguaje.
 - Buscadores, traductores, sistemas de *question answering* y sistemas industriales son...



“humanizar” la

ales que usan

s, correctores,
. y sus aplicaciones

¿Qué es el Procesamiento del Lenguaje Natural?

- El PLN es una rama de la Inteligencia Artificial orientada a “humanizar” la comunicación entre personas y máquinas.
- Está presente en prácticamente todos los productos digitales que usan lenguaje.
 - Buscadores, traductores, GPSs, sugerencias en editores, correctores, sistemas de *question answering*, minería de opiniones... y sus aplicaciones industriales son enormes.
 - El lenguaje es el mejor ejemplo de *Big Data*. Las tres *uves*: volumen, velocidad y variedad.

¿Por qué es difícil?



I'm a huge metal fan!

¿Por qué es difícil?



I'm a huge metal fan!

¿Por qué es difícil?



I'm a huge metal fan!

El PLN no es un problema uniforme

Successful NLP:
“Will a computer program ever be able to convert a piece of English text into a programmer friendly data structure that describes the meaning of the natural language text? Unfortunately, **no consensus has emerged about the form or the existence of such a data structure**”(Collobert et al., 2011).

El PLN no es un problema uniforme

- **Se puede dividir en tareas *core* (¿primarias?) y *aplicaciones*.**

Las tareas core son aquellas que en el 99% de los casos son necesarias para hacer cosas más “inteligentes”. Etiquetado morfológico, parsing sintáctico, detección de entidades nombradas, desambiguación, parsing semántico...

... y no está muy claro dónde parar ...

hasta llegar a aplicaciones complejas, como traducción, resumen, análisis de sentimiento, etc.

Successful NLP:
“Will a computer program ever be able to convert a piece of English text into a programmer friendly data structure that describes the meaning of the natural language text? Unfortunately, **no consensus has emerged about the form or the existence of such a data structure**” (Collobert et al., 2011).

El PLN no es un problema uniforme

- **El PLN es cada vez más divertido**
- Estamos ante probablemente la época más emocionante
 - Cada vez hay más datos
 - Cada vez hay más tutoriales
 - Cada vez hay más gente haciendo PLN

Successful NLP:
“Will a computer program ever be able to convert a piece of English text into a programmer friendly data structure that describes the meaning of the natural language text? Unfortunately, **no consensus has emerged about the form or the existence of such a data structure**”(Collobert et al., 2011).

El PLN no es un problema uniforme

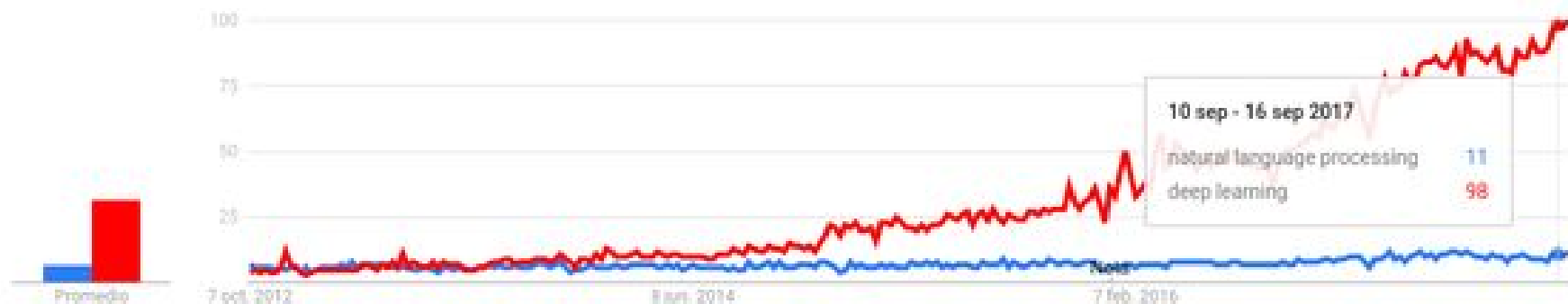
- **El PLN es cada vez más divertido**
- Estamos ante probablemente la época más emocionante
 - Cada vez hay más datos
 - Cada vez hay más tutoriales
 - Cada vez hay más gente haciendo PLN
 - ¡Cada vez hay más gente jugando a videojuegos!

Successful NLP:
“Will a computer program ever be able to convert a piece of English text into a programmer friendly data structure that describes the meaning of the natural language text? Unfortunately, **no consensus has emerged about the form or the existence of such a data structure**”(Collobert et al., 2011).

El PLN no es un problema uniforme

Successful NLP:
"Will a computer

Interés a lo largo del tiempo ?



"(Collobert et al.,
2011).

El PLN no es un problema uniforme

· Core NLP Tasks

- * Morfología
- * Sintaxis
- * Semántica
- * Reconocimiento de entidades
- * Resolución de correferencia
- * Desambiguación de sentidos (WSD)

Successful NLP:
“Will a computer program ever be able to convert a piece of English text into a programmer friendly data structure that describes the meaning of the natural language text? Unfortunately, **no consensus has emerged about the form or the existence of such a data structure**” (Collobert et al., 2011).

Problemas fundamentales en PLN

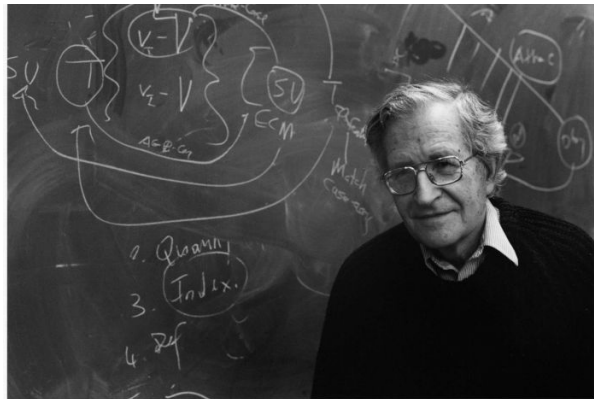
- Etiquetado morfológico o *Part-of-Speech tagging*

Problemas fundamentales en PLN

- Etiquetado morfológico o *Part-of-Speech tagging*
- **Análisis sintáctico, *parsing***
 - Análisis de constituyentes (*phrase structure grammar*)

Problemas fundamentales en PLN

- Etiquetado morfológico o *Part-of-Speech tagging*



ing

tes (*phrase structure grammar*)

Problemas fundamentales en PLN

- Etiquetado morfológico o *Part-of-Speech tagging*
- Análisis sintáctico, *parsing*
 - Análisis de constituyentes (*phrase structure grammar*)
 - Gramática de dependencias
 - Teoría de sentido texto

Problemas fundamentales en PLN

- Etiquetado morfológico o *Part-of-Speech Tagging*
- Análisis sintáctico, *parsing*
 - Análisis de constituyentes (Constituent Parsing)
 - Gramática de dependencias (Dependency Grammar)
 - Teoría de sentido texto



Problemas fundamentales en PLN

- Etiquetado morfológico o *Part-of-Speech tagging*
- **Análisis sintáctico, *parsing***
 - Análisis de constituyentes (*phrase structure grammar*)
 - Gramática de dependencias
 - Teoría de sentido texto
 - Dependencias universales!

Problemas fundamentales en PLN

- Etiquetado morfológico o *Part-of-Speech tagging*

- Análisis sintáctico, *parsing*

- Análisis de constituyentes (*phrase structure grammar*)
- Gramática de dependencias
 - Teoría de sentido texto
- Dependencias universales!



Bar-Ilan University



UPPSALA
UNIVERSITET



STANFORD



Turun yliopisto
University of Turk



UNIVERSITY OF
CAMBRIDGE



CHARLES UNIVERSITY



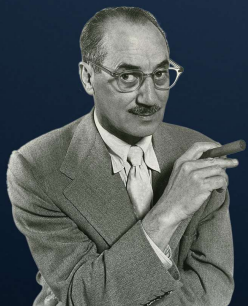
Problemas fundamentales en PLN

- Etiquetado morfológico o *Part-of-Speech tagging*
- Análisis sintáctico, *parsing*
 - Análisis de constituyentes (*phrase structure grammar*)
 - Gramática de dependencias
 - Teoría de sentido texto
 - Dependencias universales!
- Un ejemplo famoso de ambigüedad sintáctica...

Problemas f

- Etiquetado morfológico
- Link
- Análisis sintáctico,
 - Análisis de construcciones
 - Gramática de dependencias
 - Teoría de sentencias
 - Dependencias
- Un ejemplo famoso

One morning I shot an
elephant in my
pajamas. How he got
into my pajamas I'll
never know.



Groucho Marx
American Comedian
1890 - 1977

QUOTEHD.COM

Problemas fundamentales en PLN

- Reconocimiento de entidades nombradas (Named Entity Recognition)

Problemas fundamentales en PLN

- Reconocimiento de entidades nombradas (Named Entity Recognition)



Problemas fundamentales en PLN

- Reconocimiento de Entidades Nombradas (Named Entity Recognition)

• Manfred Mann's Earth Band is a British progressive rock group formed in 1971 by Manfred Mann, a keyboard player born in South Africa best known as a founding member and namesake of 60s group Manfred Mann.

Band
Music Genre
Artist
Country

Problemas fundamentales en PLN

- Resolución de coreferencia

Problemas fundamentales en PLN

- Resolución de coreferencia

- "Aunque todos vieron **al Presidente**, ninguno **lo** reconoció"
 - El pronombre "**lo**" **está en relación anafórica con el sustantivo "presidente"** (se refiere a él sin volver a nombrarlo), y **ambos ("lo" y "presidente") son correferenciales**, puesto que apuntan al mismo ente del mundo real (la persona que ejerce el cargo de presidente).
 - <https://es.wikipedia.org/wiki/Correferencia>

Problemas fundamentales en PLN

- Resolución de coreferencia

- "Aunque todos vieron **al Presidente**, ninguno **lo** reconoció"
 - El pronombre "**lo**" **está en relación anafórica con el sustantivo "presidente"** (se refiere a él sin volver a nombrarlo), y **ambos ("lo" y "presidente") son correferenciales**, puesto que apuntan al mismo ente del mundo real (la persona que ejerce el cargo de presidente).
 - <https://es.wikipedia.org/wiki/Correferencia>
- ¿Cómo se hace?
 - Coincidencias de núcleos sintácticos (Mariano Rajoy ... Rajoy).
 - Género, número, posición en frases sucesivas, lista de desambiguaciones...

Problemas fundamentales en PLN

- Desambiguación de sentidos
 - Asignar a conceptos su acepción correcta, dependiendo del contexto en que aparezcan.
 - “Esta **alianza** me ha salido más cara de lo que esperaba”

Problemas fundamentales en PLN

- Desambiguación de sentidos
 - Asignar a conceptos su acepción correcta, dependiendo del contexto en que aparezcan.
 - “Esta **alianza** me ha salido más cara de lo que esperaba”



Problemas fundamentales en PLN

- Desambiguación de sentidos
 - Asignar a conceptos su acepción correcta, dependiendo del contexto en que aparezcan.
 - “Esta **alianza** me ha salido más cara de lo que esperaba”



Problemas fundamentales en PLN

- Desambiguación de sentidos
 - Asignar a conceptos su acepción correcta, dependiendo del contexto en que aparezcan.
 - “Esta **alianza** me ha salido más cara de lo que esperaba”



Aplicaciones

- Resumen
- Simplificación
- Análisis de sentimiento
- Respuestas a preguntas
- Semántica
- Traducción automática
- ...

“Every time I fire a linguist out of my group, the accuracy goes up”
Bedřich Jelínek (1932, 2010), IBM Research & John Hopkins University

¡Gracias!

¡Gracias!



@luisanke