

No ~~habrá~~ hay PLN sin lingüística

jueves, 21 de diciembre de 2017 | Campus Madrid



Javier G. Sogo

 @jgsogo

I. Presentación

II. Contexto

III. Pesimismo

IV. Una nueva esperanza

V. El filólogo estructurado

VI. Maridaje




Lingwars
@lingwars

Siguiendo






Hola!

9:05 - 15 nov. 2017



Javier G. Sogo
BBVA | Quant developer
@jgsogo





lingwars
@lingwars

- **Lengua + tecnología (ars/artis)**
- **Conferencias**
- **Meetups sobre PLN**
- **Talleres para lingus**
- **GAPLEN (NLTK)**

- Rayadas varias (Twitter)
- Cerveza (3er tiempo)

••
w
@lingwars



**MAP SHOWING
THE ROUTE OF COLUMBUS
ON HIS FIRST VOYAGE ACROSS
THE ATLANTIC AND RETURN
TO ILLUSTRATE HIS UTILIZATION
OF THE WINDS AND CURRENTS**

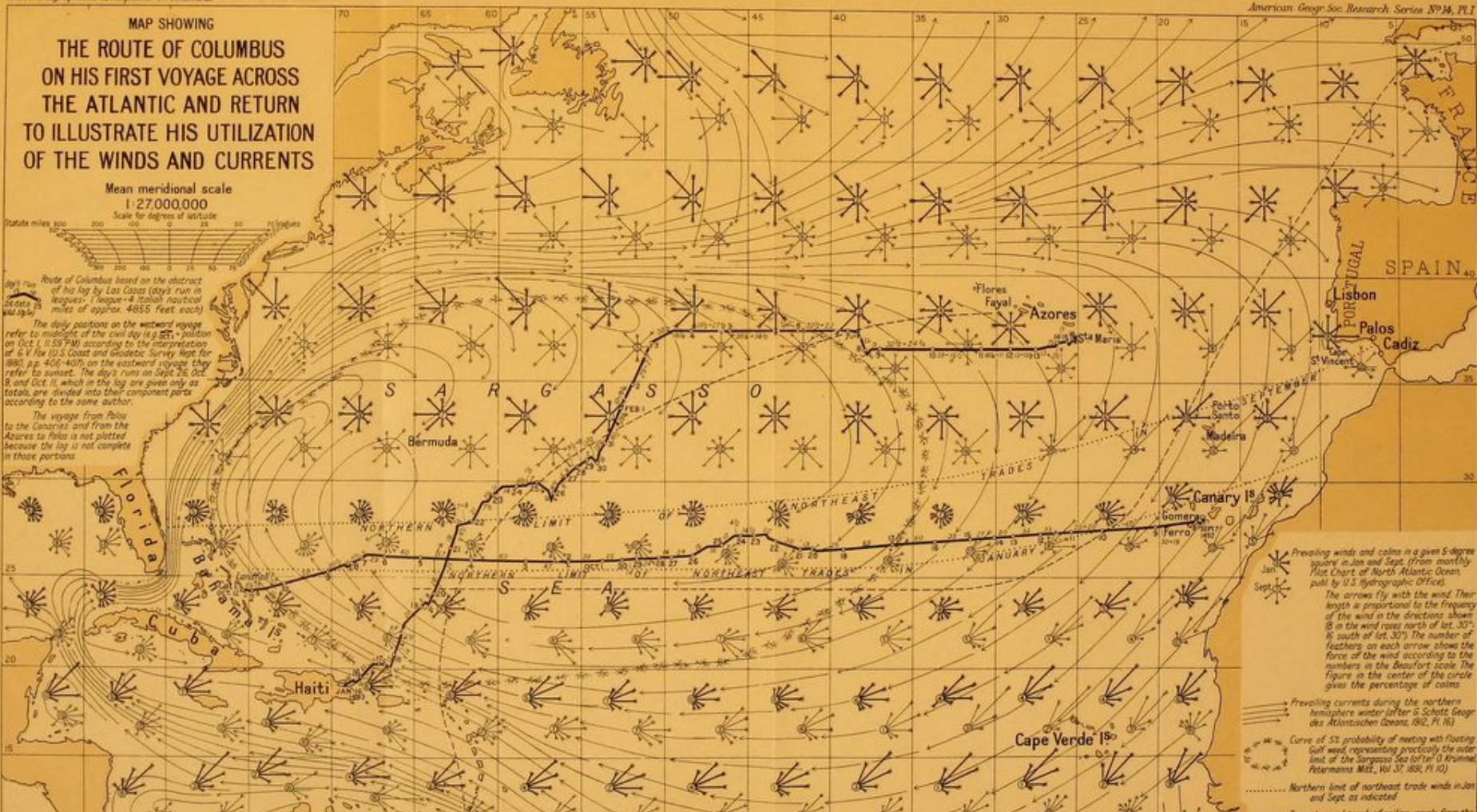
Mean meridional scale
1:27,000,000
Scale for degrees of altitude



Day's Run
The route of Columbus based on the abstract of his log by Las Casas (day's run in leagues; 1 league = 4 Italian nautical miles of approx. 4855 feet each)

The daily positions on the westward voyage refer to midnight of the civil day (i.e. GMT) on Oct. 1, 11:59 PM) according to the interpretation of G. V. Fox (U.S. Coast and Geodetic Survey Rept. for 1880, p. 406-407) on the eastward voyage they refer to sunset. The day's runs on Sept. 25, Oct. 9 and Oct. 11, which in the log are given only as totals, are divided into their component parts according to the same author.

The voyage from Palos to the Canaries and from the Canaries to Palos is not plotted because the log is not complete in those portions.



Prevailing winds and calms in a given S. degree every 5 min and Sept. (from monthly Weather Chart of North Atlantic Ocean, pub. by U.S. Hydrographic Office).

The arrows fly with the wind. Their length is proportional to the frequency of the wind in the directions shown. In the wind roses north of lat. 30° N. south of lat. 30° S. The number of feathers on each arrow shows the force of the wind according to the numbers in the Beaufort scale. The figure in the center of the circle gives the percentage of calms.

Prevailing currents during the northern hemisphere winter (after G. Schott, Geogr. des Atlantischen Ozeans, 1912, Pl. 16)

Curve of 5% probability of meeting with floating Gulf weed, representing practically the outer limit of the Sargasso Sea (after O. Krummel, Petermanns Mit., Vol. 37, 1891, Pl. 10)

..... Northern limit of northeast trade winds in Jan and Sept. as indicated

--- Present usual track of sailing vessels from the English Channel to West Indies and from

I. Presentación

II. Contexto

III. Pesimismo

IV. Una nueva esperanza

V. El filólogo estructurado

VI. Maridaje



Jorge Diz Pico 

@xurxodiz

Siguiendo



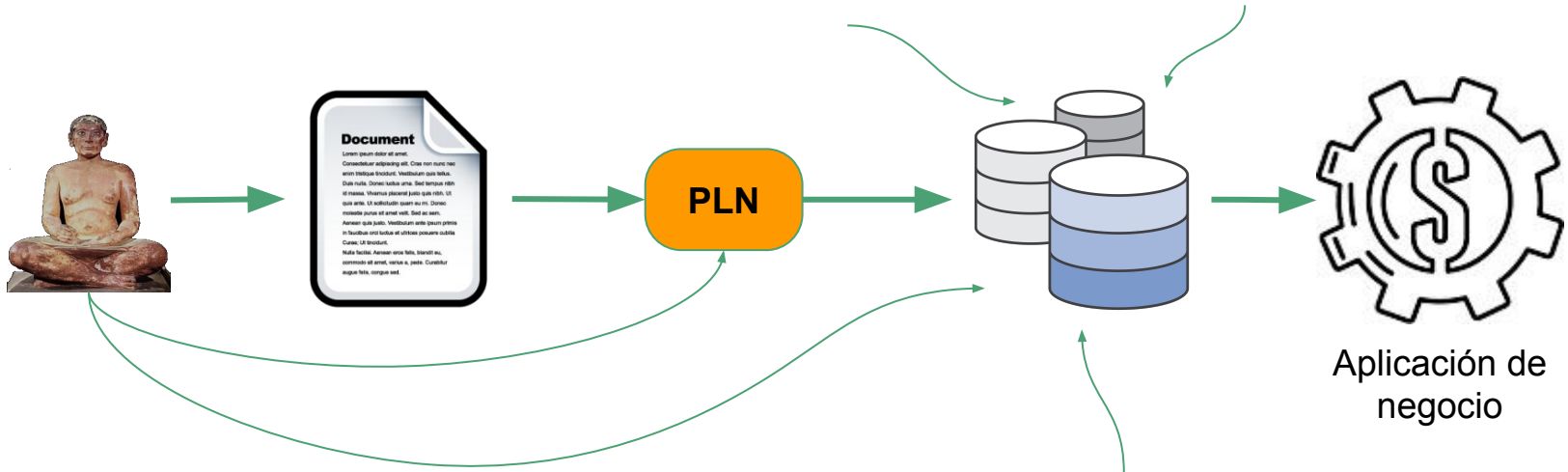
But languages are far too gargantuan, complicated non-entities to be abstracted into a discrete algorithm input.

 Traducir del inglés

17:50 - 13 dic. 2017

¿Qué entiendo por PLN?

Conjunto de técnicas, modelos, herramientas, algoritmos,... que nos permiten **convertir un texto expresado en lenguaje natural a datos procesables** por un algoritmo

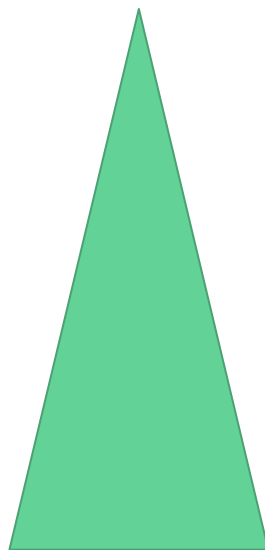


Cuáles son los retos y dónde estamos

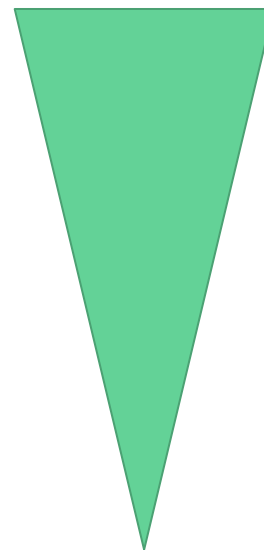
Tasks

Classification
Clustering
Summarization
Extraction
Topic modeling
Translation dialogue
Question answering
Inference speech act

Dependency on NLP



Scalability



Hongning Wang, CS@UVa “Introduction to Natural Language Processing” (dic-2017)

Tareas PLN (actualizado y completo): https://github.com/Kyubyong/nlp_tasks

I. Presentación

II. Contexto

III. Pesimismo

IV. Una nueva esperanza

V. El filólogo estructurado

VI. Maridaje



François Chollet

@fchollet

Siguiendo

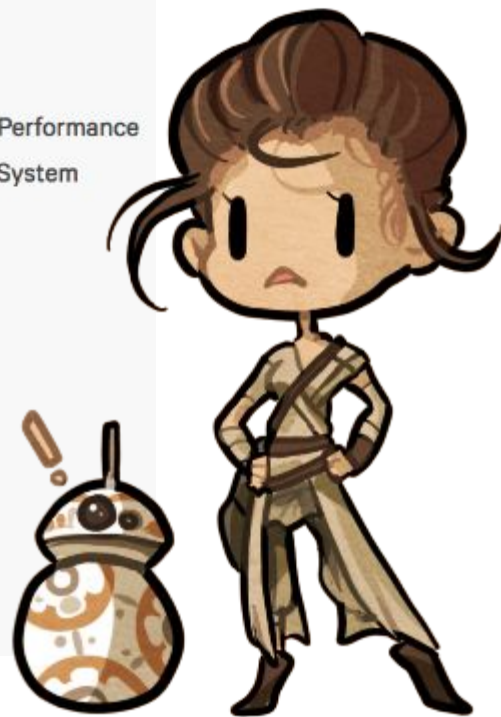
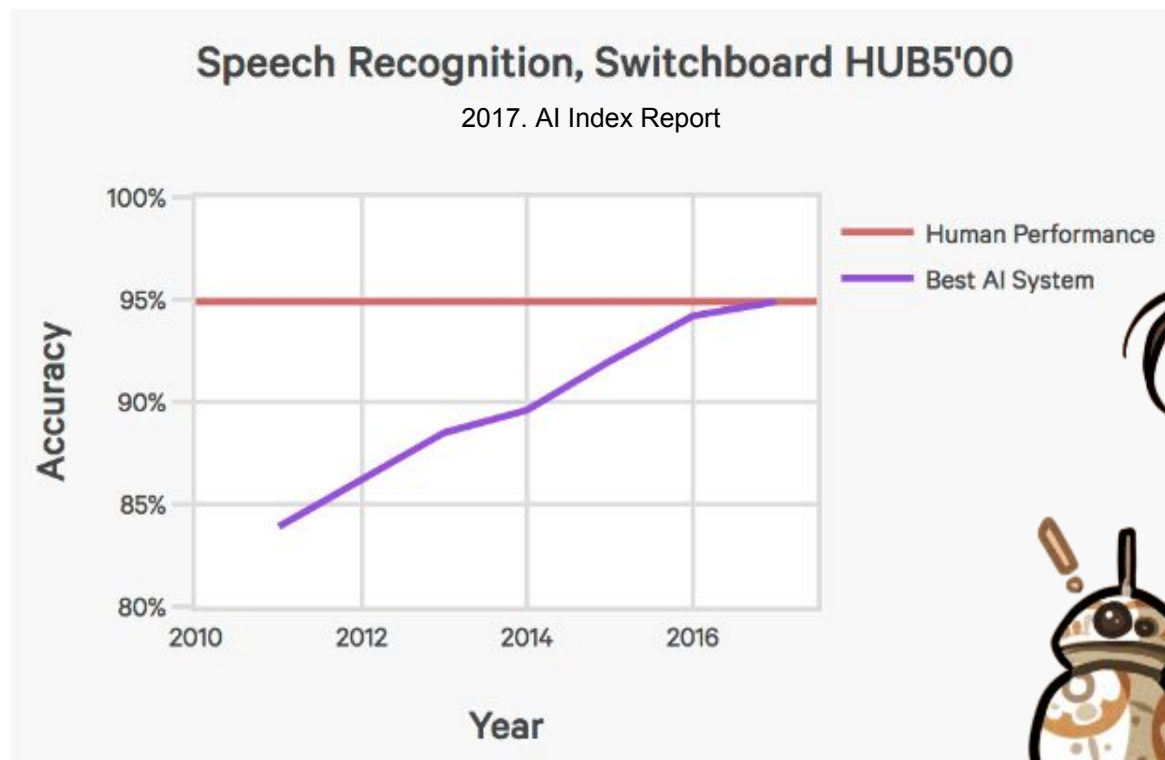


For all the progress made, it seems like almost all important questions in AI remain unanswered. Many have not even been properly asked yet.

Traducir del inglés

7:32 - 2 mar. 2017

Chatbots, ¿por qué no #Talkbots?



Cómo codificar el lenguaje

- ¿Ontologías? (RDF, OWL)
- ¿Corpus?
- ¿Vectores?

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. Until such fundamental

Collobert *et al.* 2011. Natural Language Processing (Almost) from Scratch

I. Presentación

II. Contexto

III. Pesimismo

IV. Una nueva esperanza

V. El filólogo estructurado

VI. Maridaje



Elena Álvarez Mellado

@lirondos

Siguiendo



¿Qué es la Lingüística Computacional? ¿Para qué sirve? Hoy se cumplen ¡4 años! de esta charla que tantas alegrías me ha dado



Todo lo que un filólogo no sabe que sabe hacer

El 18 de diciembre, en el Aula Magna de la Facultad de Filosofía y Letras, día. Elena Álvarez (empresa El Molino de Ideas) nos presentó la conferencia: Todo ...

[youtube.com](https://www.youtube.com)

14:35 - 18 dic. 2017

Semántica distribucional

- ¡Volveremos con una **almáciga**!

- Y que sea bonita... y que no cueste mucho

[...]

- Hasta los que venden **almácigas** están sometidos a presiones económicas en esta época de la historia.

- ¿Has dicho **almácigas**?

- Sí! yo trabajo en **almácigas**, soy **almaciguero**, me llaman el **Almaciguero** Mayor, arreglo, diseño y vendo **almácigas**.



Semántica distribucional

Word embeddings:

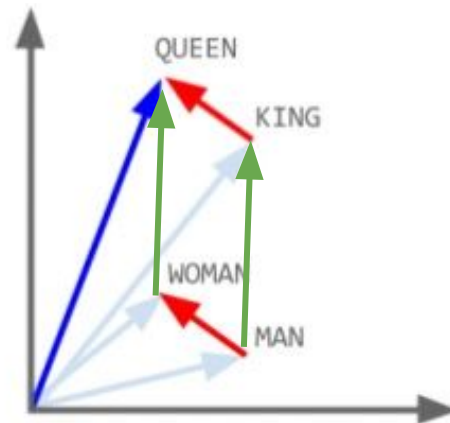
- word2vec (Mikolov et al., 2013)
- GloVe (Pennington et al., 2014)



John Rupert Firth
(1890 - 1960)
lingüista británico

“You shall know a word by
the company it keeps”
(Firth, 1957)

So king - man + woman = queen!



sexo → queen - king = woman - man

poder → queen - woman = king - man



Año 2015: Blue Origin (Bezos) ponen un cilindro en vertical en el suelo



Año 1911: Amudsen llega al polo sur geográfico

Image courtesy National Library of Australia

Semántica distribucional

Ventajas:

- Inducir significados a partir de texto **sin anotar**

Trabajo pendiente:

- Cómo aplicarlo a realizar inferencias
- Tiempo/datos de entrenamiento (vectores pre-entrenados)
- Interpretabilidad
- Ética

- I. Presentación
- II. Contexto
- III. Pesimismo
- IV. Una nueva esperanza

V. El filólogo estructurado

- VI. Maridaje



Oscar Corcho

@ocorcho

Siguiendo

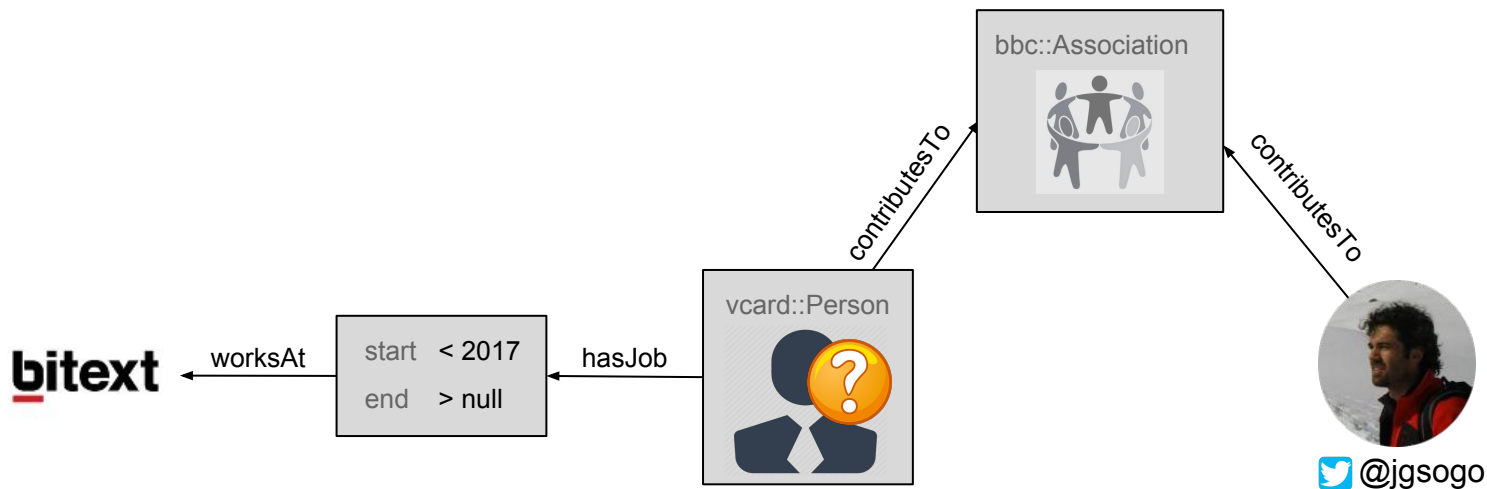


Now [@juansequeda](#) telling us [@KCAP2017](#) about how to apply semantic technologies in the real world

Ontologías - Las inferencias al alcance de la mano

Si convertimos el texto en grafos las inferencias son búsquedas “sencillas”

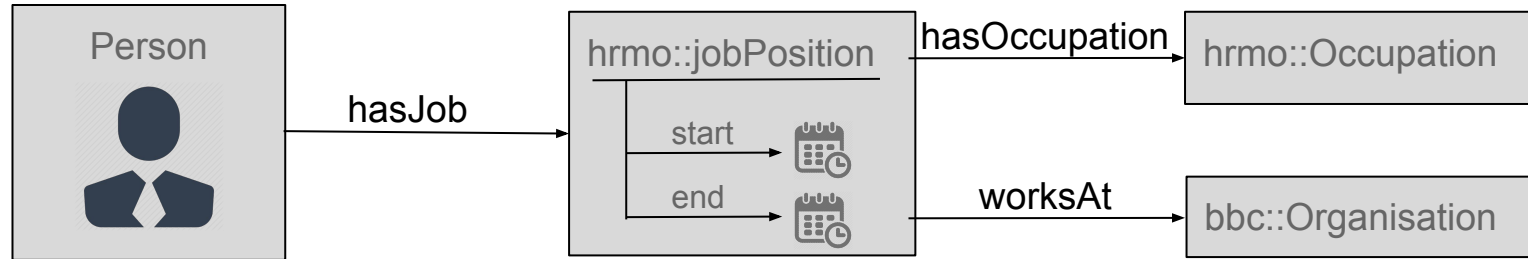
¿ Qué persona trabaja en BiText y contribuye en la misma asociación que @jgsogo ?



Ontologías - El lingüista es nuestro hombre

Trabajar

2. tener una ocupación remunerada en una empresa, una institución, etc.. (DLE)



Ontologías - La lingüista también



Leticia MFM

@nimbusaeta

Siguiendo



Cómo no le va a gustar a un filólogo programar, si es algo en lo que se pueden **RECORRER DICCIONARIOS**



20:12 - 30 nov. 2017



Una base de datos con la mente del lingüista

Ontologías

Ventajas:

- Sin problemas de ambigüedad ni interpretabilidad

Inconvenientes:

- Centradas en los sustantivos
- Dependientes del idioma
- Enfoque prescriptivo
- Coste de actualización **prohibitivo**
- ¿Sesgos?

- I. Presentación
- II. Contexto
- III. Pesimismo
- IV. Una nueva esperanza
- V. El filólogo estructurado

VI. Maridaje



Luis

@luisanke

Siguiendo



I have written a blog post encouraging humanists to join [#nlproc](#), with a few ideas on where your expertise can fit in.

Combinar enfoque simbólico y estadístico

-- Manos a la obra --

Centrarse en un único dominio (domótica, legal, restaurantes,...) e idioma

Construir una ontología

Embeddings para expandir la ontología (IA no supervisada)

Corpus de preguntas para entrenar

PLN pipeline: POS, lemmatizar,...

Grafo → Motor de inferencias

#Talkbot



Combinar enfoque simbólico y estadístico

-- Investigación --

- Módulo de lenguaje controlado
- **Interpretabilidad:** eliminar comportamientos indeseados
- **Ética:** sesgos
- **Legalidad:** auditoría de modelos

“[Salvo consentimiento explícito o autorización legal específicos,] todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar”

Art. 22 - Reglamento (UE) 2016 / 679 - Protección de Datos Personales



Muchas gracias




Javier G. Sogo

 @jgsogo



Lingwars

 @lingwars