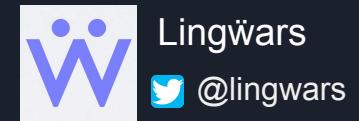


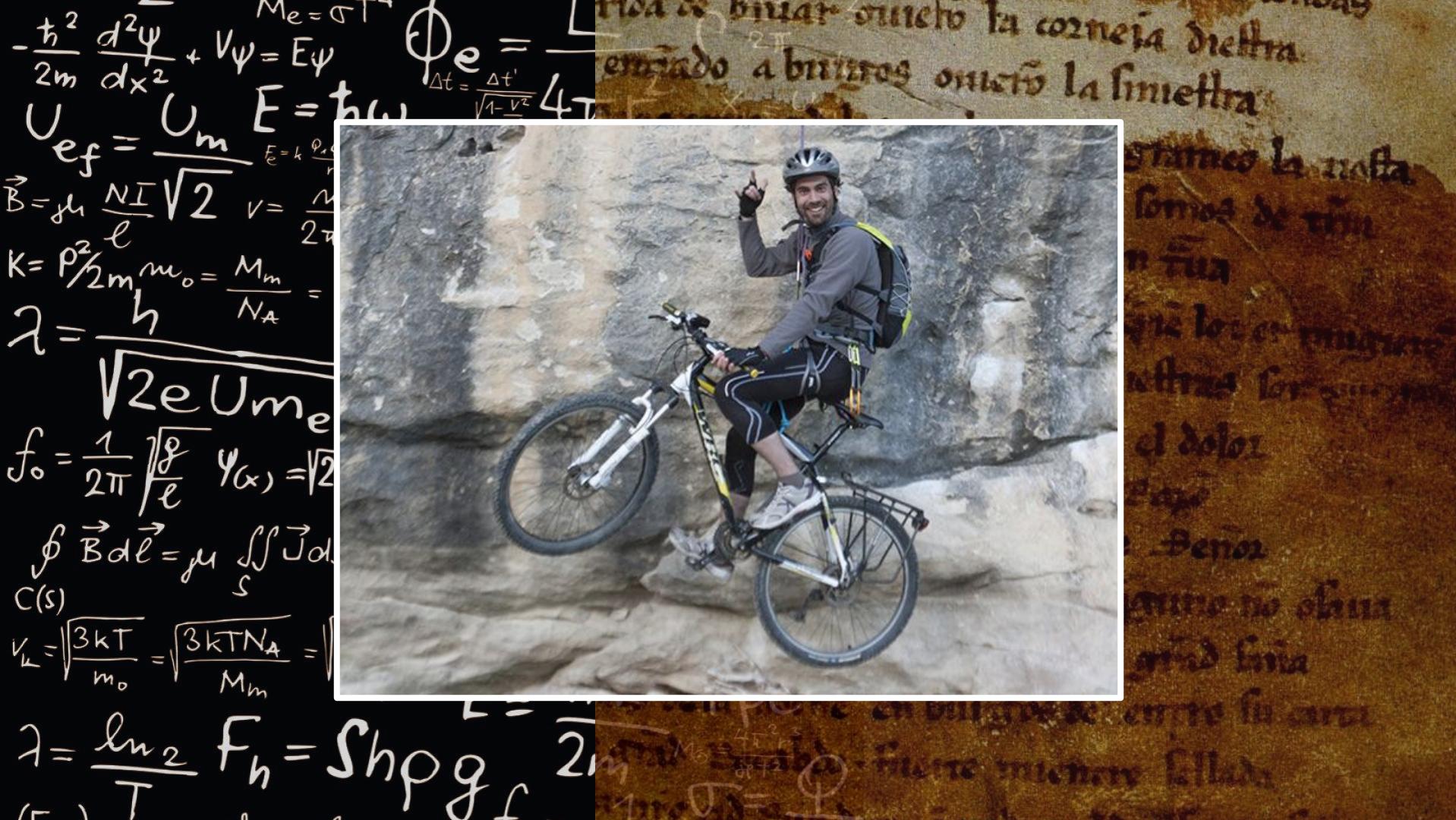


# No habrá PLN sin ontologías

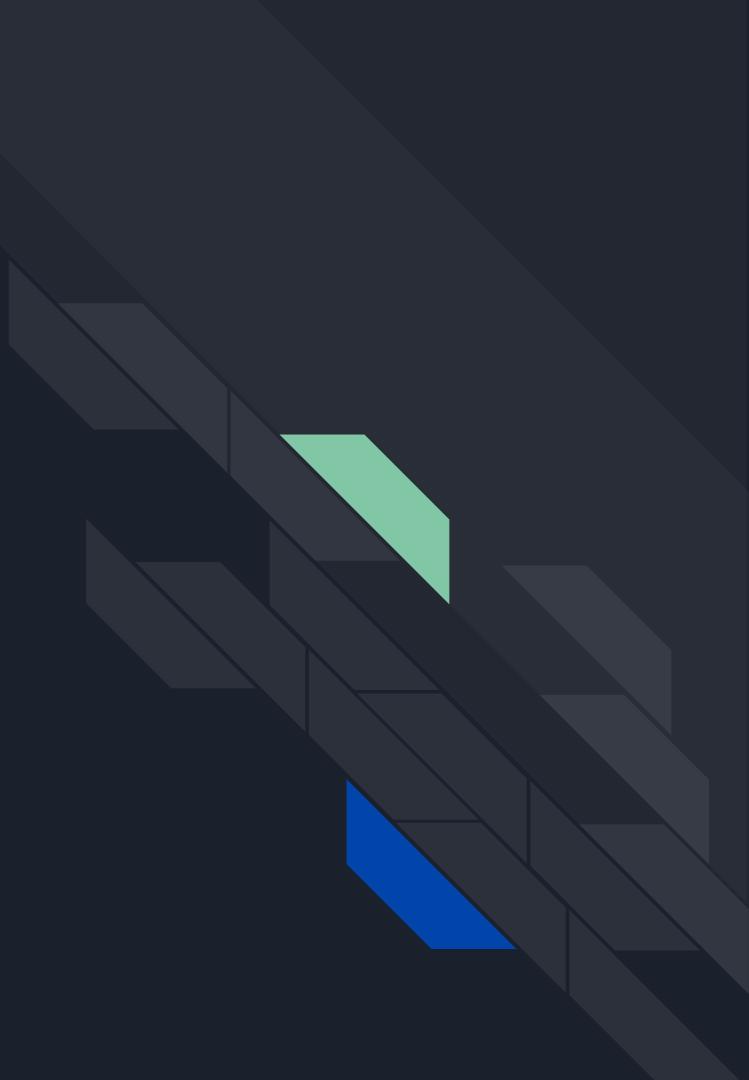


Javier G. Sogo  
 @jgsogo

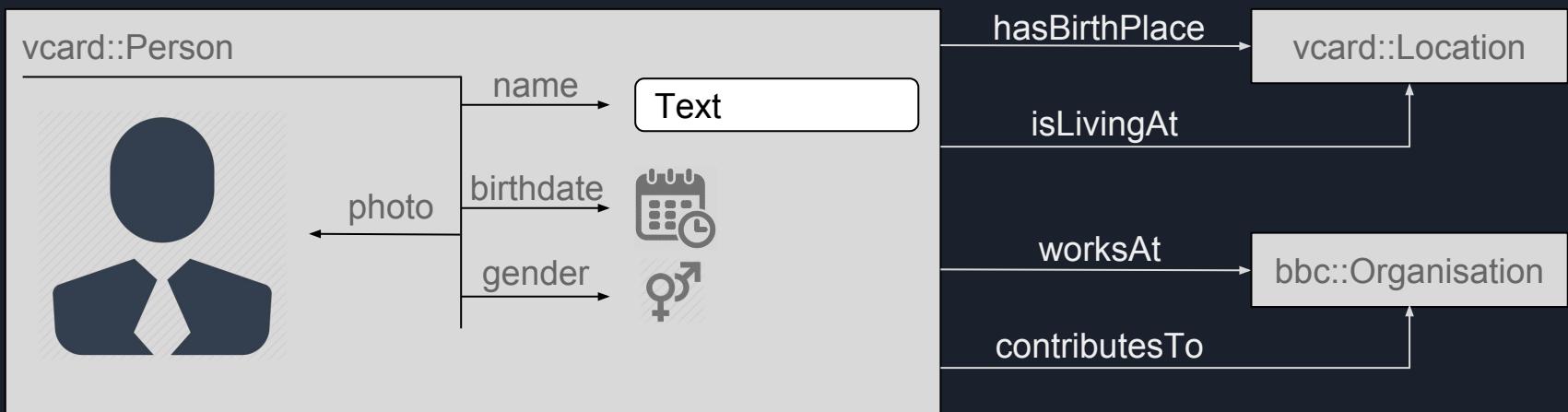




¿Quién es @jgsogo?

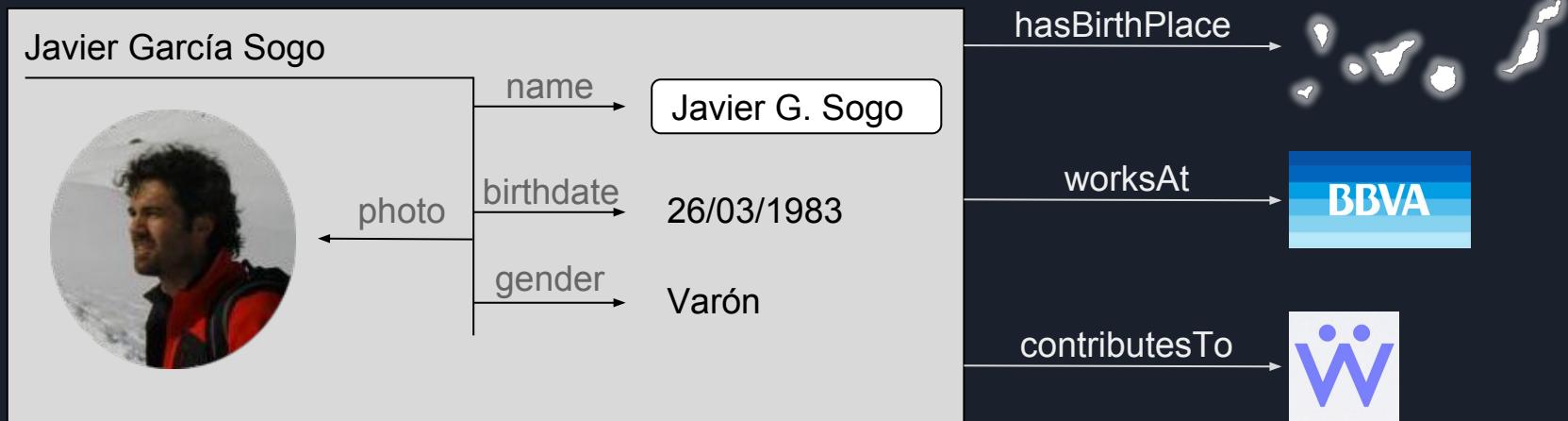


# ¿Quién es @jgsogo?

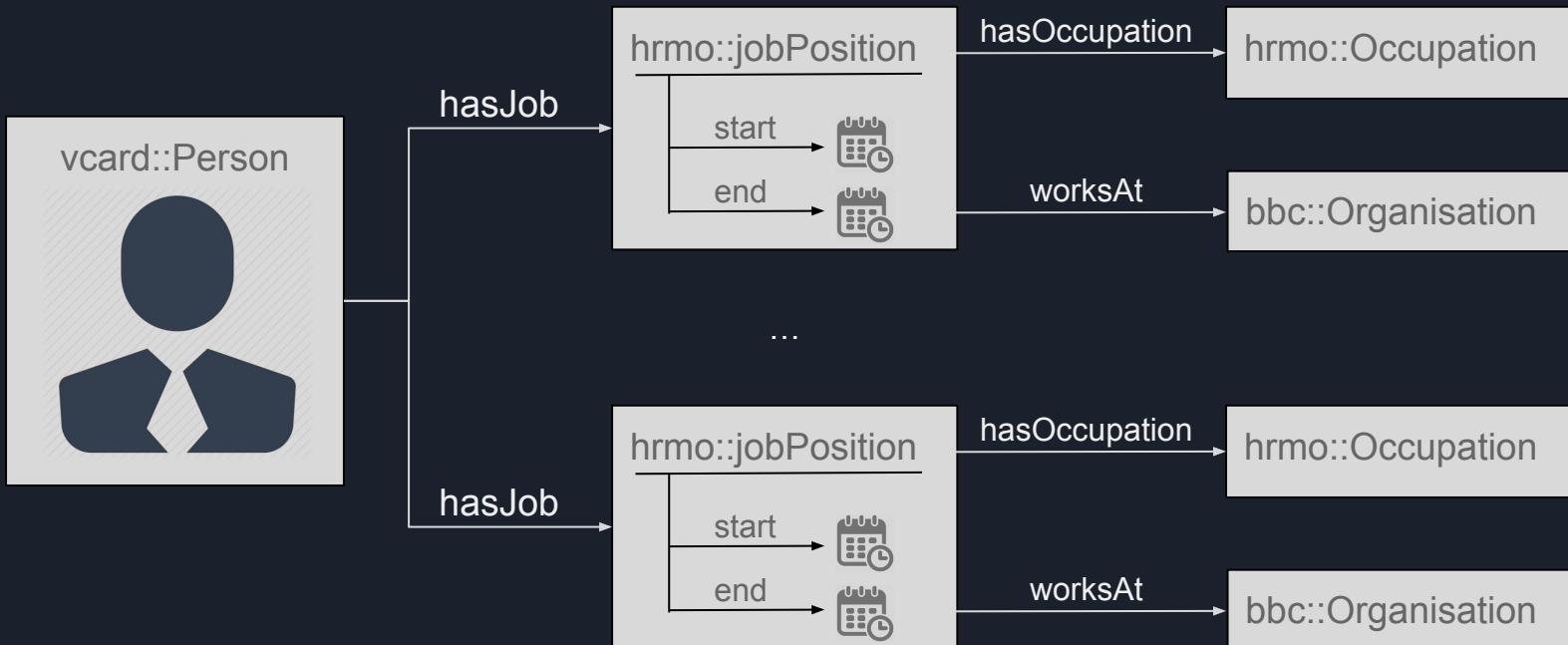


vcard: vCard Ontology (<https://www.w3.org/TR/vcard-rdf/>) mantenida por la W3C  
bbc: BBC Ontology (<https://www.bbc.co.uk/ontologies>) mantenida por la BBC

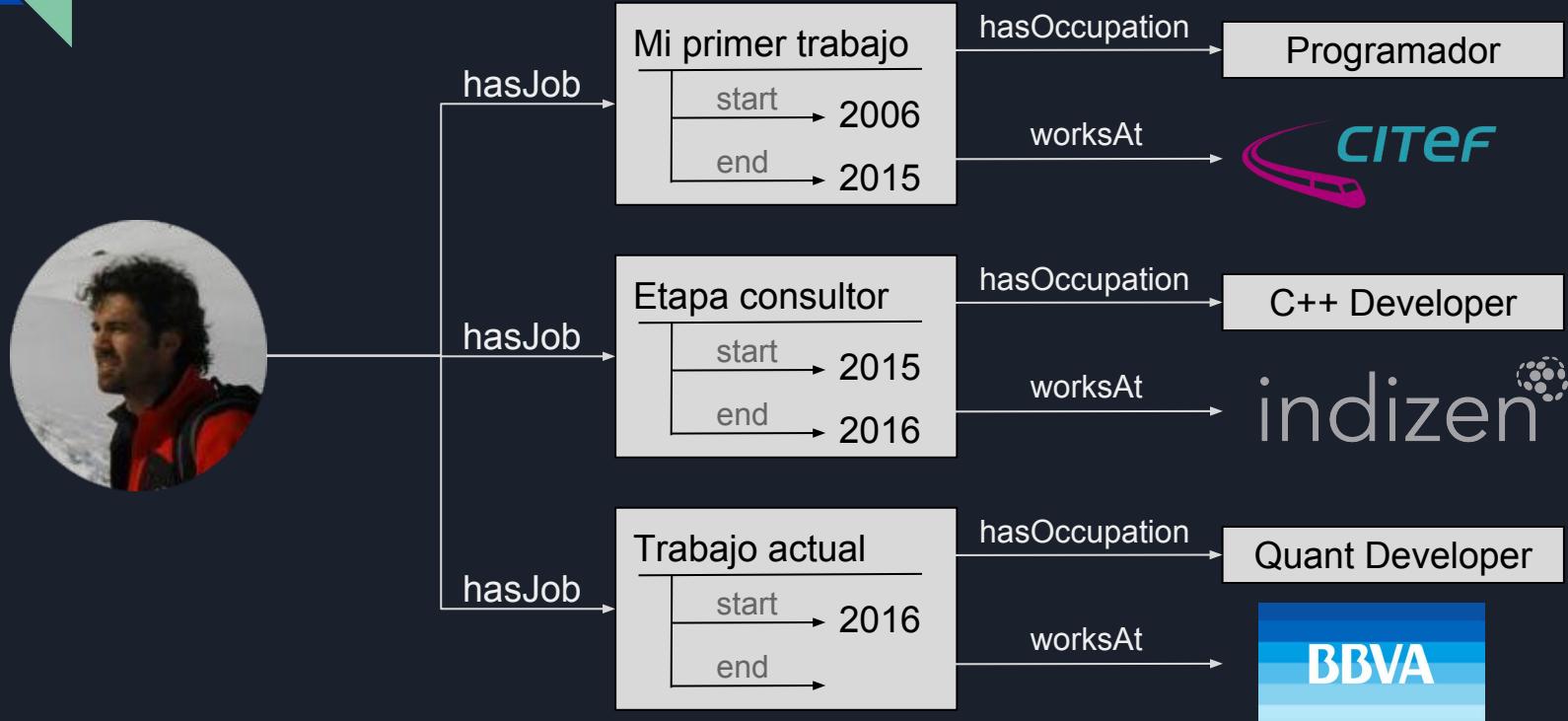
# ¿Quién es @jgsogo?



# ¿Quién es @jgsogo?



# ¿Quién es @jgsogo?



# ¿Y eso de Lingwars?

Grupo de personas (“asociación”)

Objetivo:

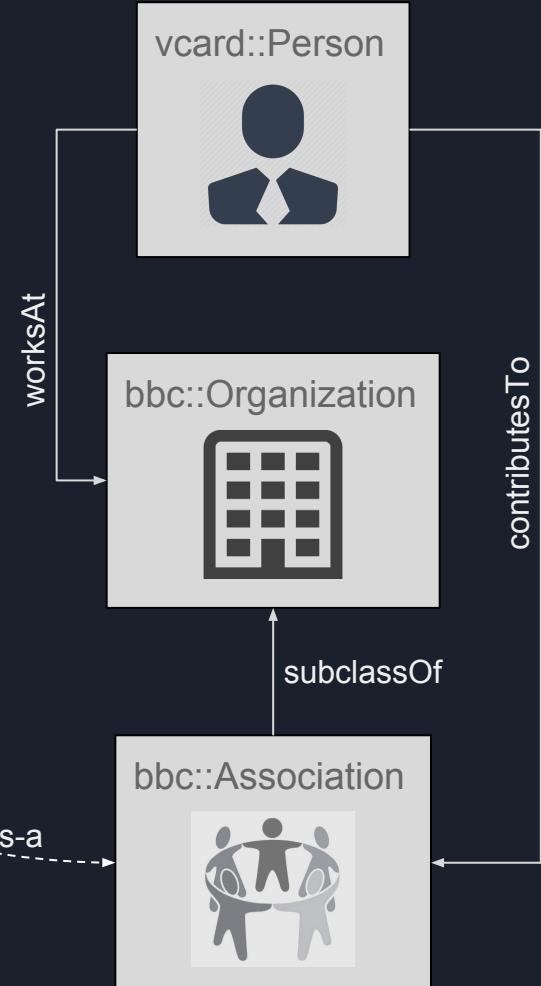
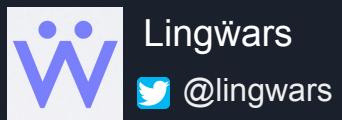
- derribar el muro entre ciencias y letras

Somos:

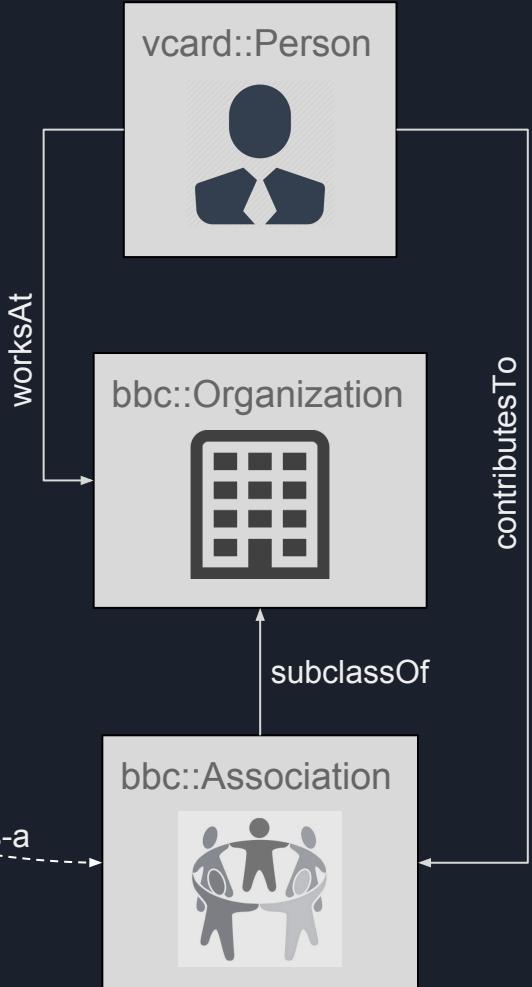
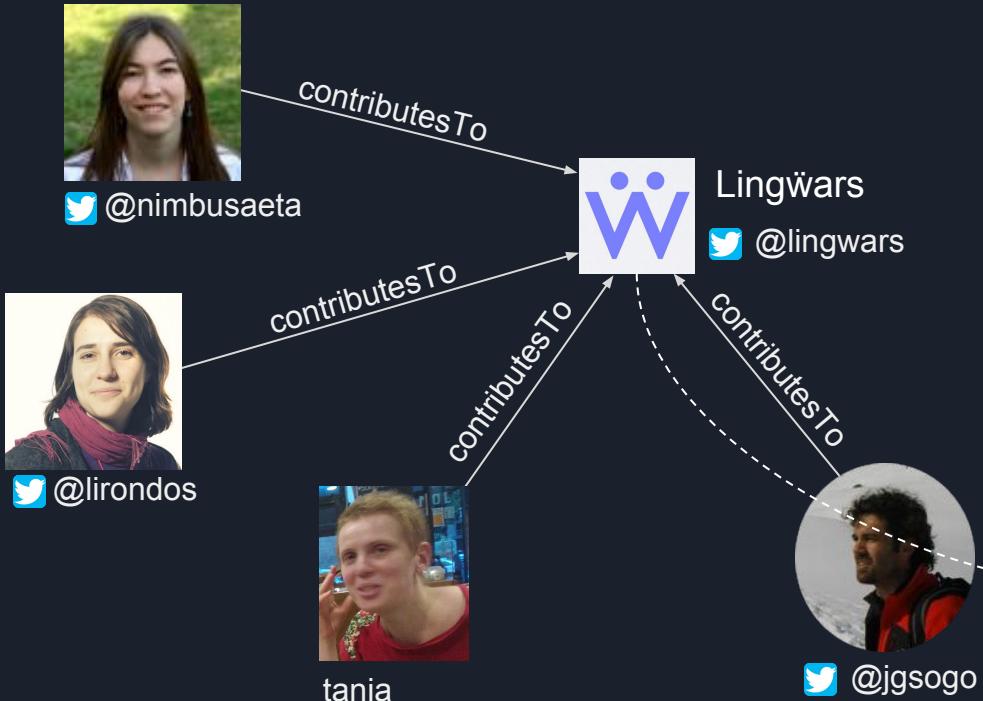
- techies humanistas
- lingüinis filotecnólogos

Organizamos:

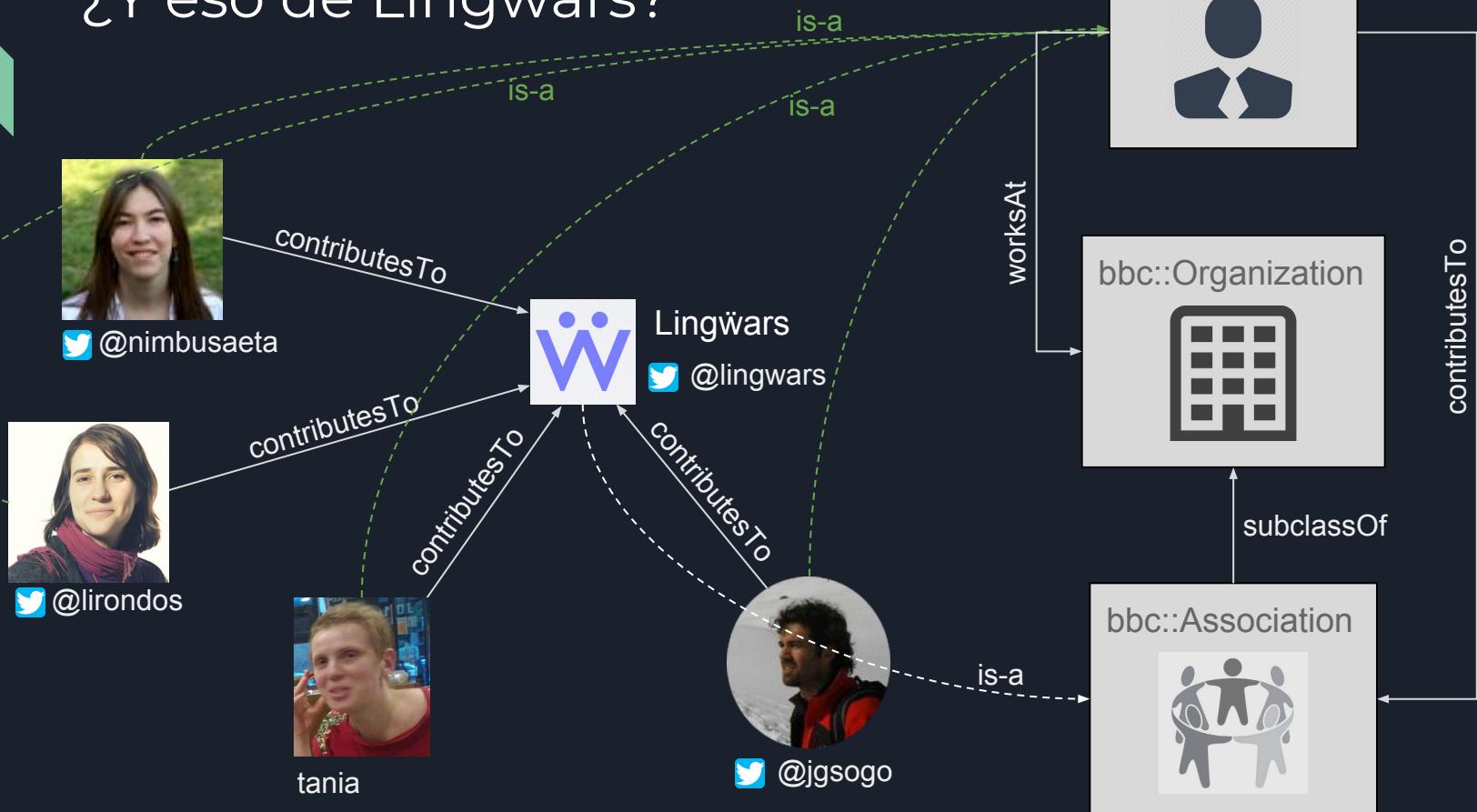
- eventos
- talleres
- charlas



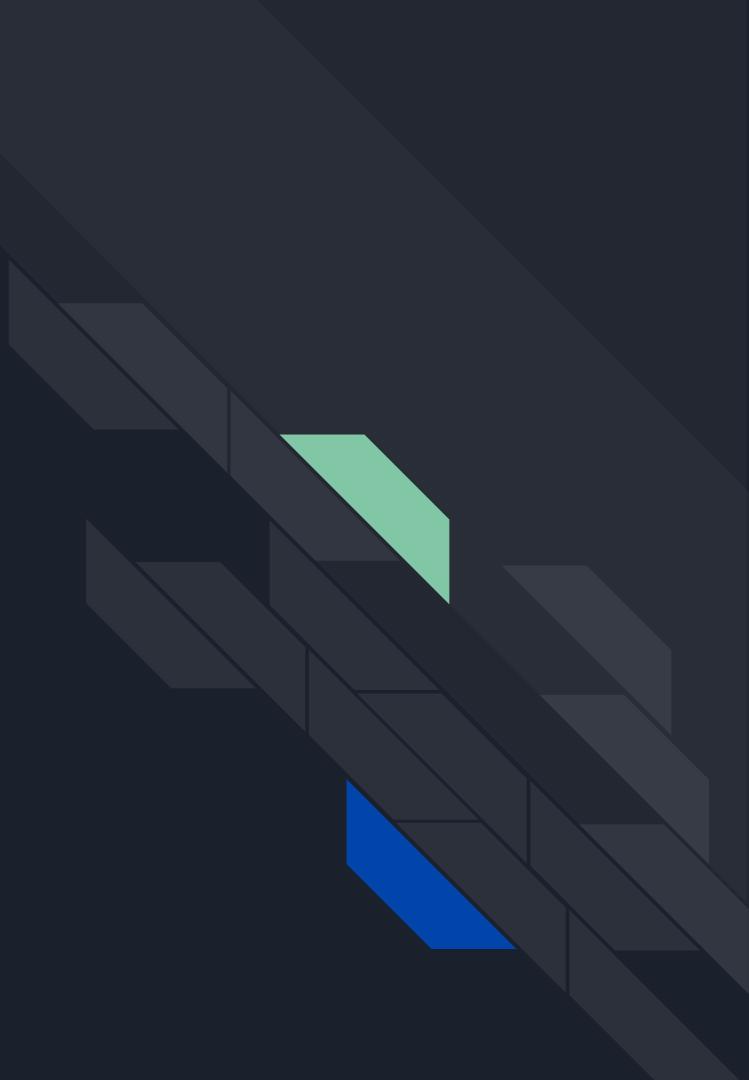
# ¿Y eso de Lingwars?



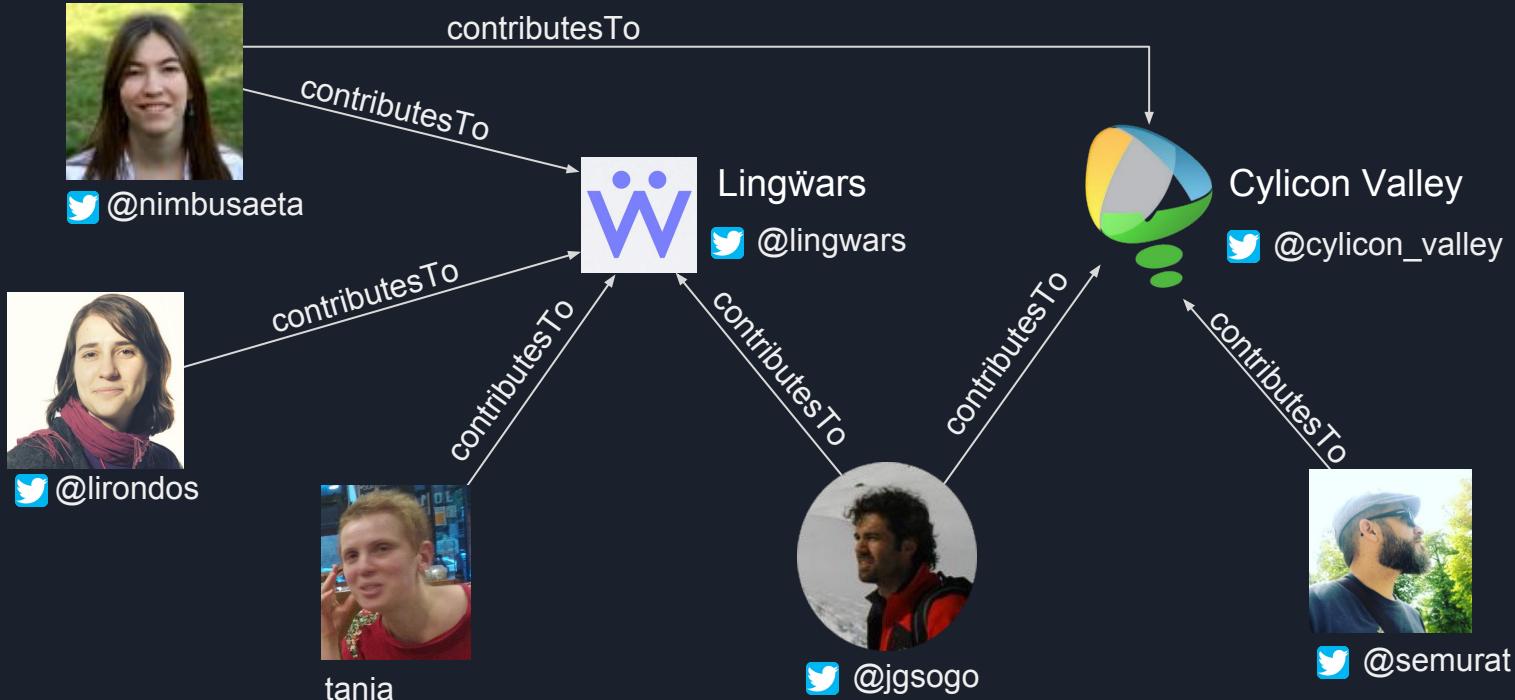
# ¿Y eso de Lingwars?



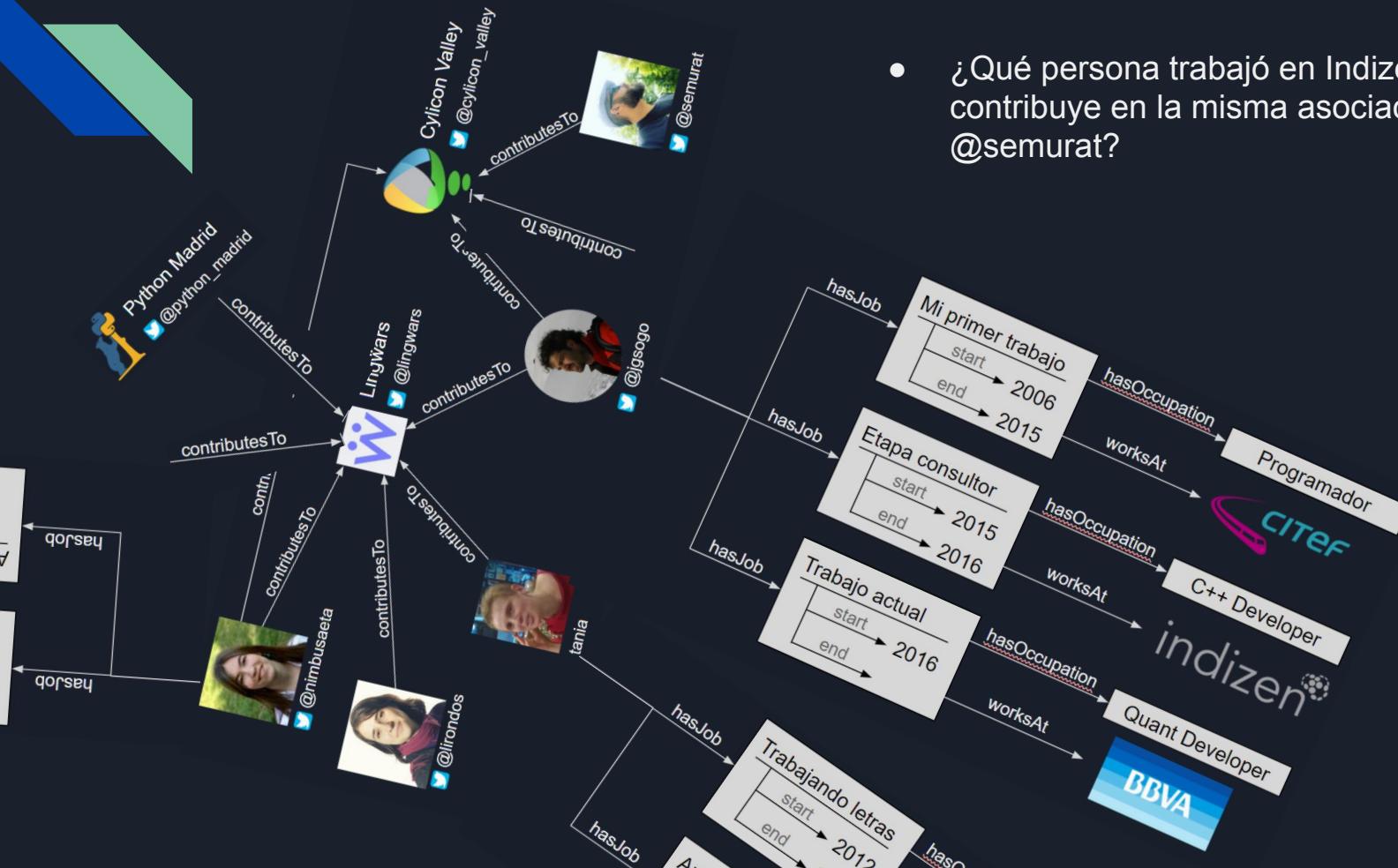
Más información sobre  
@jgsogo



# Más información



- ¿Qué persona trabajó en Indizen en 2015 y contribuye en la misma asociación que @semurat?



- ¿Qué personas actualmente están trabajando en BBVA y que contribuyan en asociaciones de Valladolid?





Internet.  
Buscando información



A cartoon illustration of Rick Sanchez from the TV show "Rick and Morty". He has his signature blue spiky hair, a white lab coat over a teal shirt, and a shocked or surprised expression with wide eyes and an open mouth. He is gesturing with his hands, palms facing up, as if he is reacting to something unexpected.

Búsqueda por “palabras clave”

10.000.0000.0000000 resultados

**Recuperación de información\* ⇒ Documentos**



# Extracción de información\* ⇒ Datos

\*Information extraction

Imagen: Papou "Shouting in despair" en DevianArt



# El objetivo final en PLN

**Convertir el lenguaje en datos estructurados**, de tal forma que una máquina sea capaz de trabajar con él:

- Buscar datos concretos
- Generar lenguaje
- Resumir contenidos

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



# El objetivo final en PLN

Utilizar Internet como una **gran base de datos global**:

- Convertir todo en datos estructurados
- Realizar inferencias sobre los datos

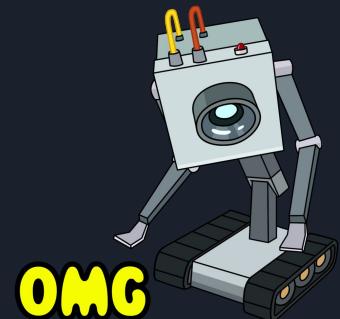
The Semantic Web is an **extension** of the current Web in which information is given **well-defined meaning**, better enabling computers and people to work in **cooperation**. It is based on the idea of having data on the Web defined and **linked** such that it can be used for more effective discovery, automation, integration, and reuse across various applications. For the

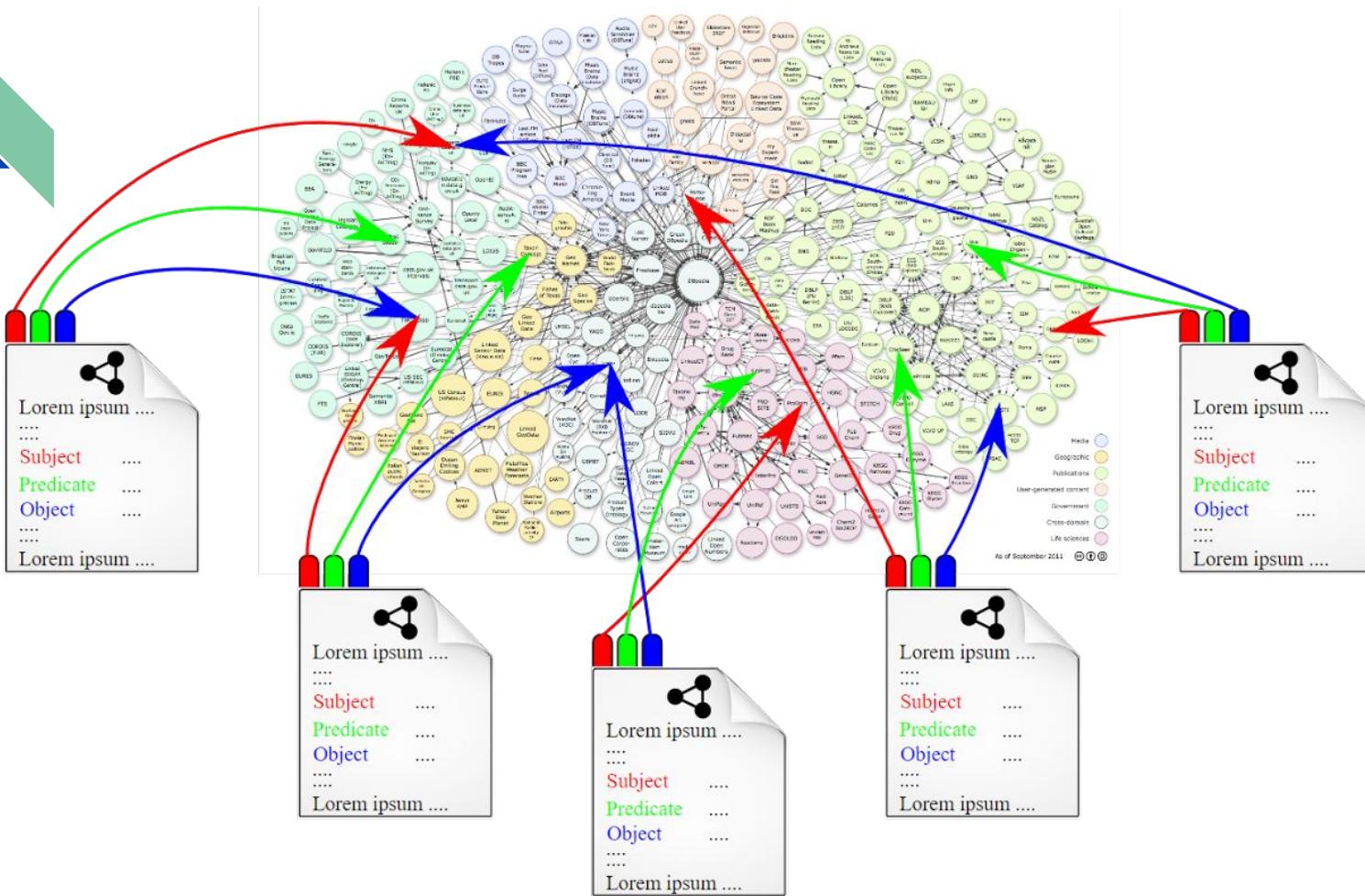
Hendler, J., Berners-Lee, T., Miller, E. Integration Applications on the Semantic Web, 2002

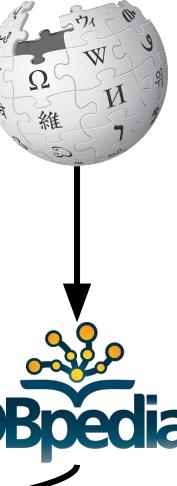
# La Web Semántica (Web 3.0)

Los datos están representados de tal forma que las máquinas pueden entenderlos:

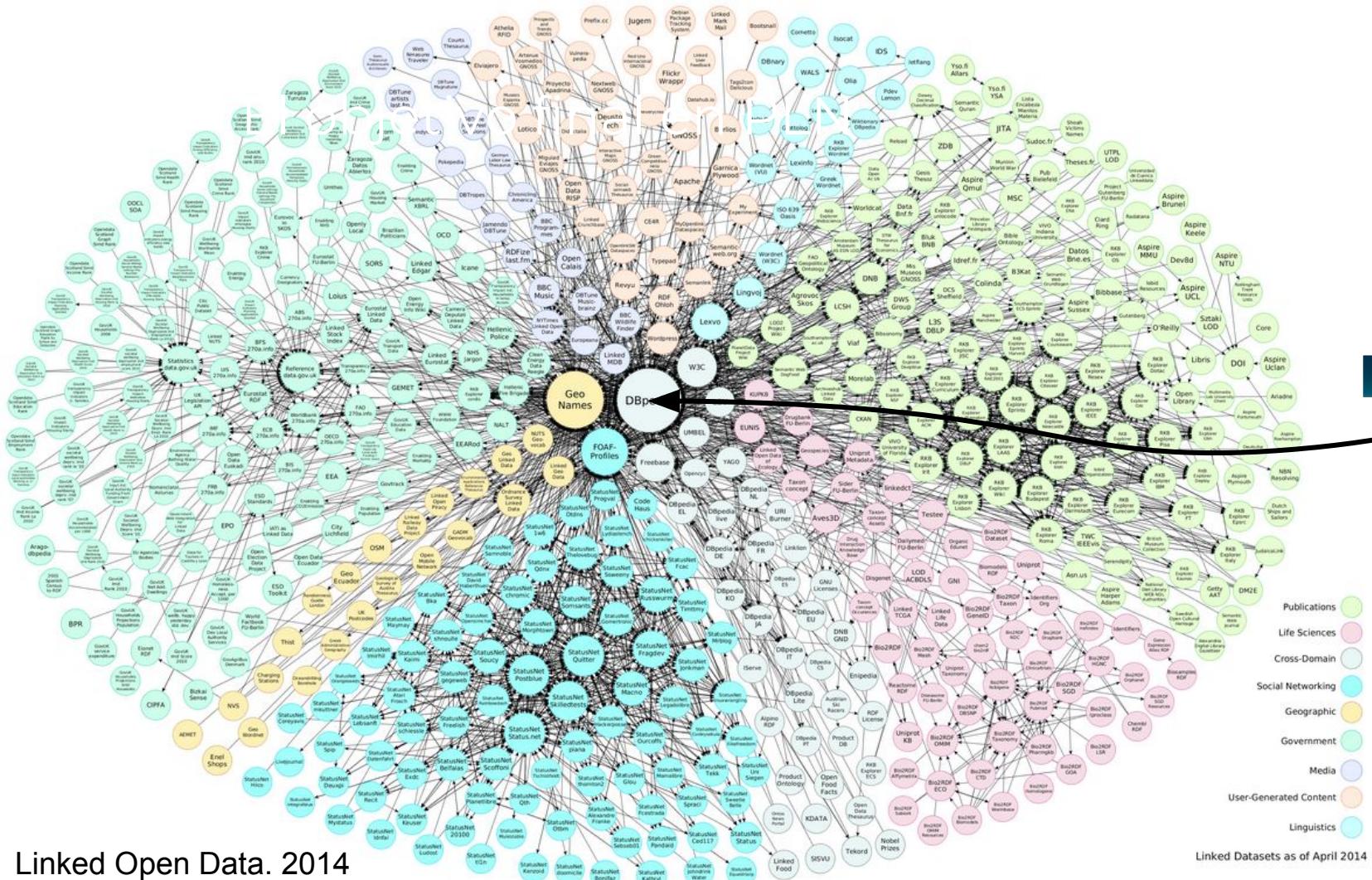
- Encontrar datos concretos en un documento
- Acceder a información relacionada con los datos
- Realizar inferencias
- Asistentes personales







**DBpedia**





Lo que piensa....

...el ingeniero





Lo que piensa....

...el lingüista



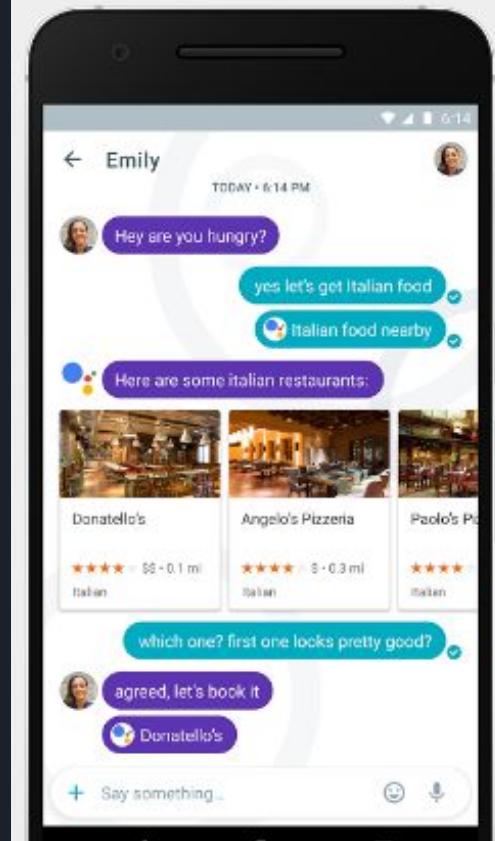


La realidad es...

... un chatbot para  
pedir comida



Get answers in your chat from  
the Google assistant



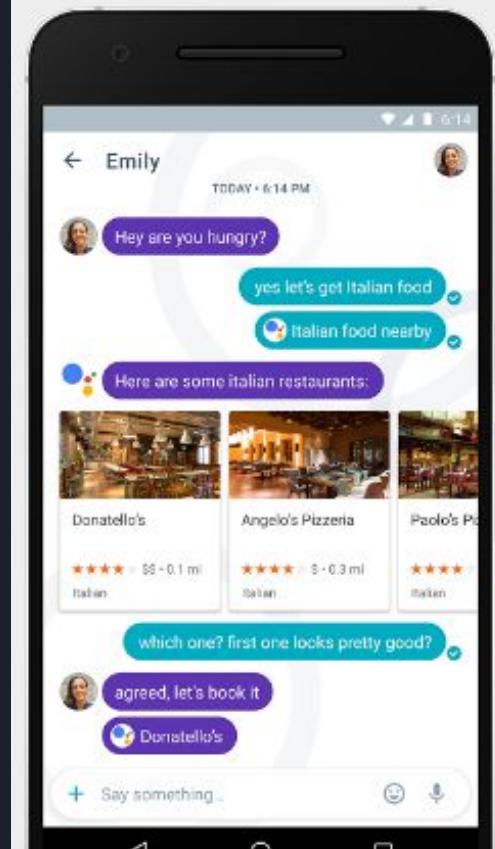


La realidad es...

... un chatbot para  
pedir ~~comida~~ pizza



Get answers in your chat from  
the Google assistant



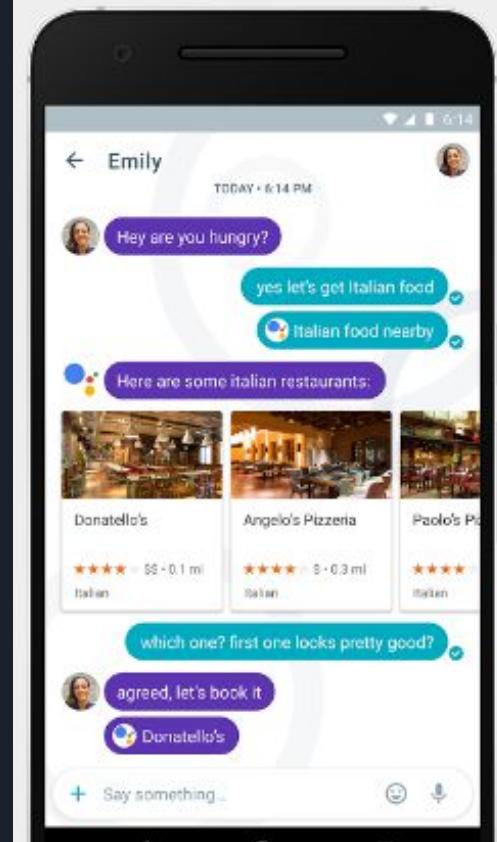


La realidad es...

... un chatbot para  
pedir ~~comida~~ pizza  
en inglés



Get answers in your chat from  
the Google assistant



Convertir el lenguaje en  
datos estructurados



# Convertir el lenguaje en datos estructurados

¿Qué persona trabajó en Indizen en 2015 y contribuye en la misma asociación que @nimbusaeta?

1. Tokenizar
2. Pos-tagging
3. Resolver correferencias
4. Reconocimiento de entidades
5. Semántica
6. Codificación de dependencias



# Paso 1. Tokenizar

Identificar oraciones y palabras

¿Qué persona trabajó en Indizen en 2015 y contribuye en la misma asociación que @nimbusaeta?

- En **español** es sencillo separar palabras. Hay muy pocas excepciones.
- ¿Y separar las oraciones? Atención a abreviaturas, acrónimos,...
- **Ambigüedad**: el contexto nos indica qué tiene sentido considerar una palabra

en | dos | mil | quince

en dos mil quince

en | dos mil quince

## Paso 2. Etiquetar pos-tagging

Rol gramatical (Part Of Speech) de cada palabra

¿	Qué	persona	trabajó	en	Indizen	en	2015	y	contribuye	en	la	misma	asociación	que	@nimbusaeta	?
DET	NOUN	VERB		NOUN		NUM		VERB					NOUN			NOUN

- Análisis morfológico de toda la vida.
- ¿Qué conjunto de etiquetas utilizar?
  - EAGLES ⇒ castellano
  - Universal dependencies ⇒ conjunto válido para todas las lenguas
- **Ambigüedades.**

# Paso 3. Resolver correferencias

Identificar términos que hacen referencia al mismo ente

¿Qué persona trabajó en Indizen en 2015 y contribuye en la misma asociación que @nimbusaeta?

DET

NOUN

VERB

NOUN

NUM

VERB

NOUN

NOUN



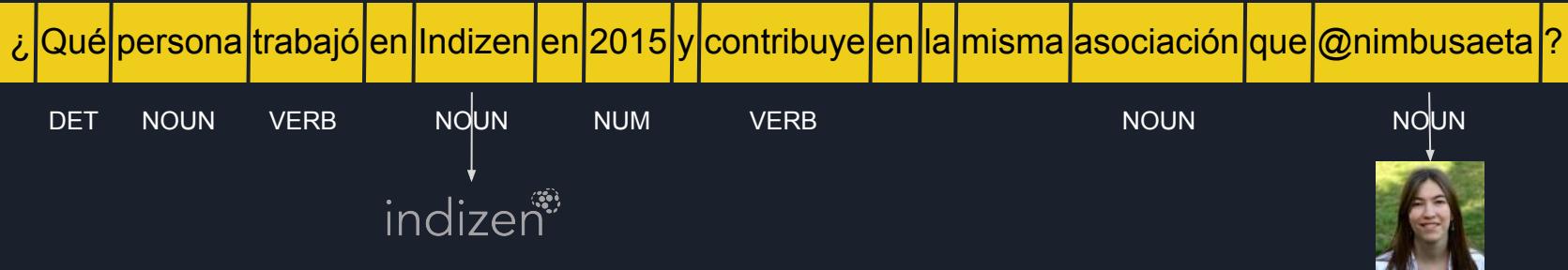
One morning I shot an elephant in my pajamas.  
How he got into my pajamas I'll never know.

(Groucho Marx)

izquotes.com

## Paso 4. Identificar entidades nombradas

Named Entity Recognition (NER): personas, lugares, organizaciones,...



- Personas: nombres propios, nicks,...
- Mayúsculas (español),...
- Búsqueda en bases de datos (DBpedia)

# Paso 5. Desambiguar semánticamente

Utilizar el contexto para elegir el significado

¿Qué persona trabajó en Indizen en 2015 y contribuye en la misma asociación que @nimbusaeta?

DET

NOUN

VERB

NOUN

NUM

VERB

NOUN

NOUN

indizen<sup>®</sup>

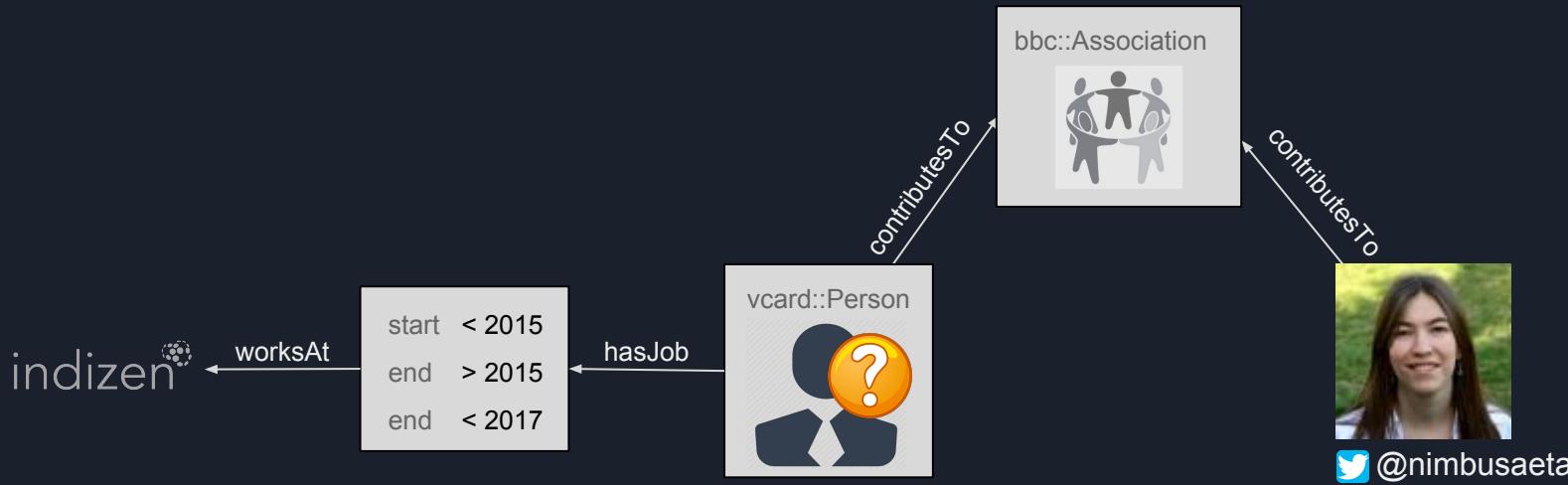


- “Encontrar en el diccionario” la entrada correspondiente a cada palabra (lematizar!).
- Persona:
  - 1. f. Individuo de la especie humana
  - 8. f. Gram. Categoría gramatical inherente en algunos pronombres [...]

# Paso 6. Codificación en una estructura

Codificar el lenguaje en una estructura inteligible para la máquina

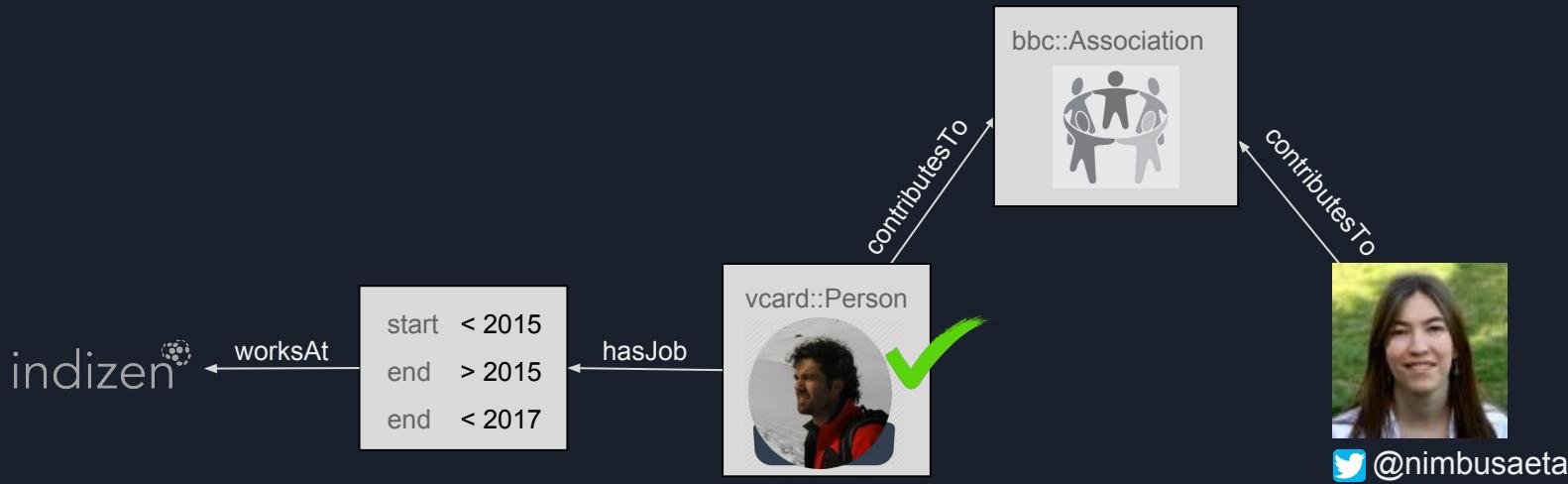
¿Qué persona trabajó en Indizen en 2015 y contribuye en la misma asociación que @nimbusaeta?



# Paso 6. Codificación en una estructura

Codificar el lenguaje en una estructura inteligible para la máquina

¿Qué persona trabajó en Indizen en 2015 y contribuye en la misma asociación que @nimbusaeta?



# Ontologías



# Inteligencia artificial

Fracaso tras fracaso

1315. Raimundo Lulio. Razonamiento mecánico

1950. Alan Turing. Juego de imitación

Redes neuronales

1960 - 1974. Sistemas basados en reglas -- Problemas de búsqueda

Representación del conocimiento

Procesamiento de lenguaje natural (Guerra Fría)

1980 - 1987. Sistemas expertos

Redes neuronales (de verdad)

1993 - 2001. Algoritmia

2001. Internet: Big Data, Deep learning, buzz words....





# Ontología. Definición

Gruber (1993). Explicit specification of a conceptualization.

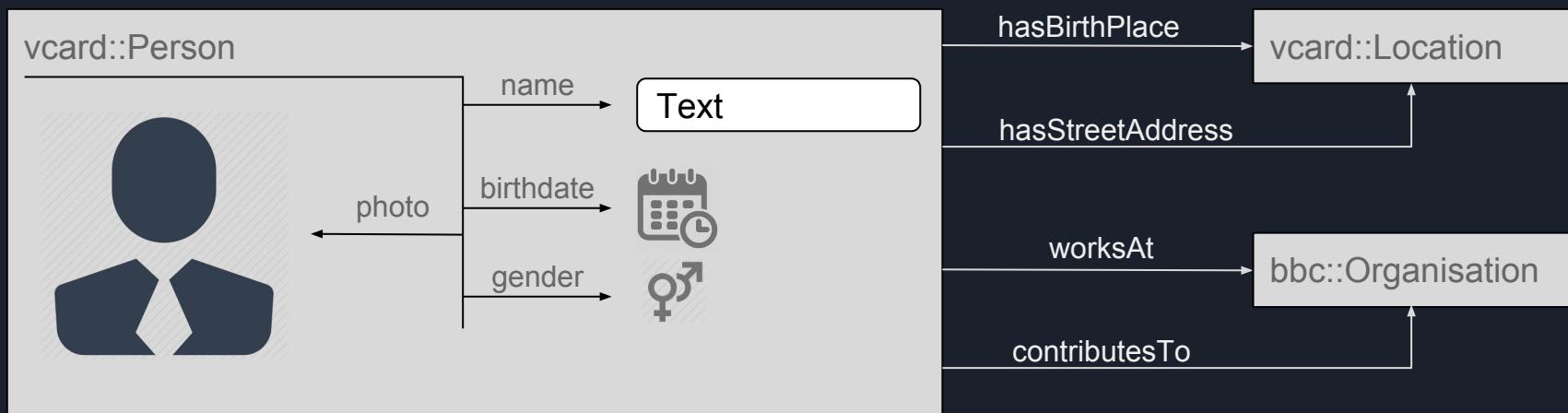
Borst (1997). Formal specification of a shared conceptualization.

Studer (1998). An ontology is a formal, explicit specification of a shared conceptualization.

Domain	Ceñida a un dominio de conocimiento
Explícita	Sin ambigüedades
Formal	La máquina tiene que poder entenderlo
Compartida	Tiene que haber consenso

# Ontología. Componentes

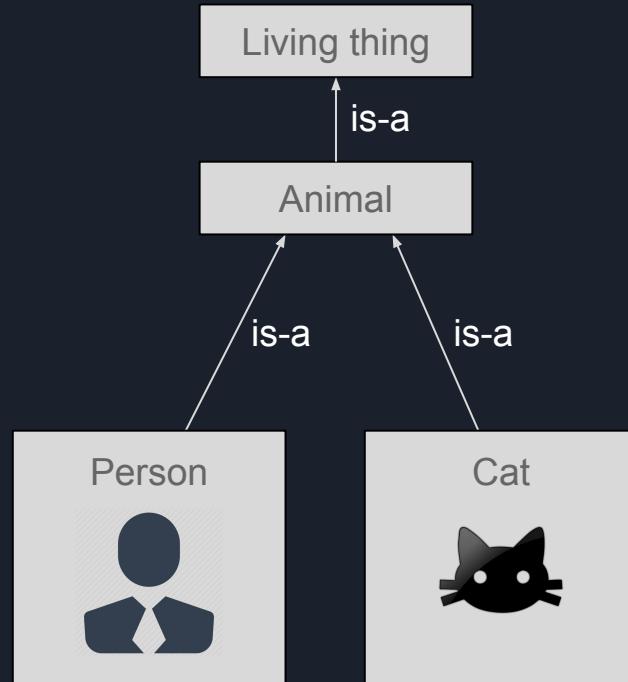
1. Clases - Conceptos
2. Atributos - Propiedades (de un concepto)
3. Relaciones (entre conceptos)



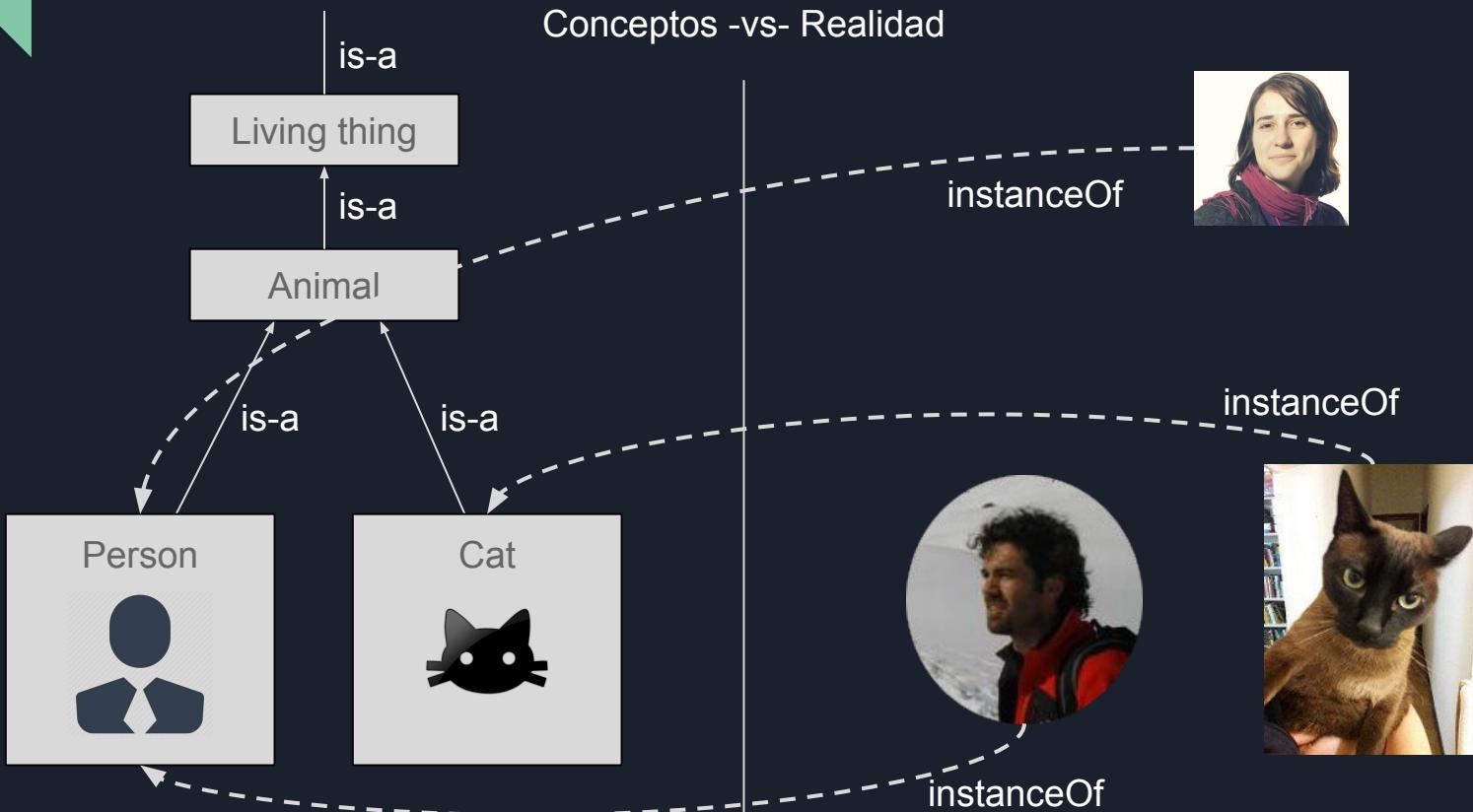
# Ontología. Conceptos

Son los nodos de la red:

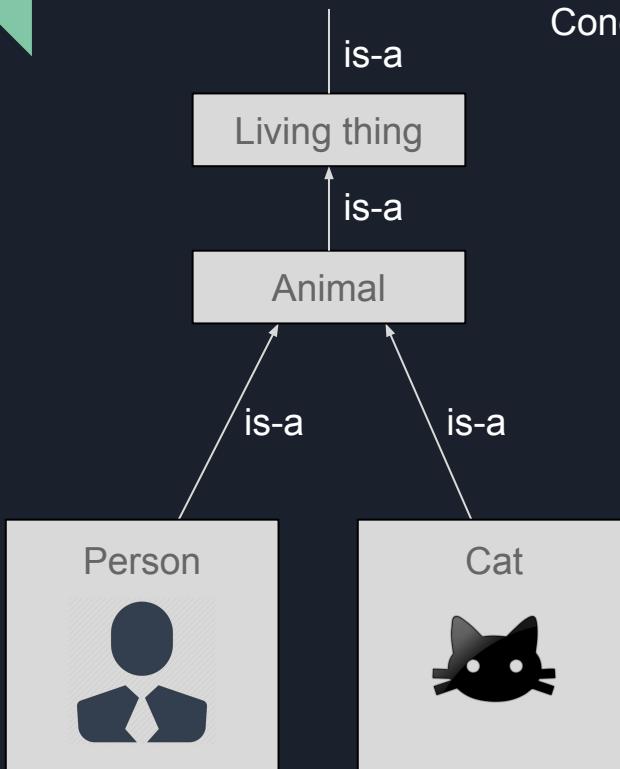
- Se definen por sí mismos (o con atributos)
- Se organizan en taxonomías
  - relación is-a: hiperonimia/hiponimia
- Propiedades:
  - completitud
  - conocimiento disjunto



# Ontología. Clases - Conceptos



# Ontología. Clases - Conceptos



Conceptos -vs- Realidad

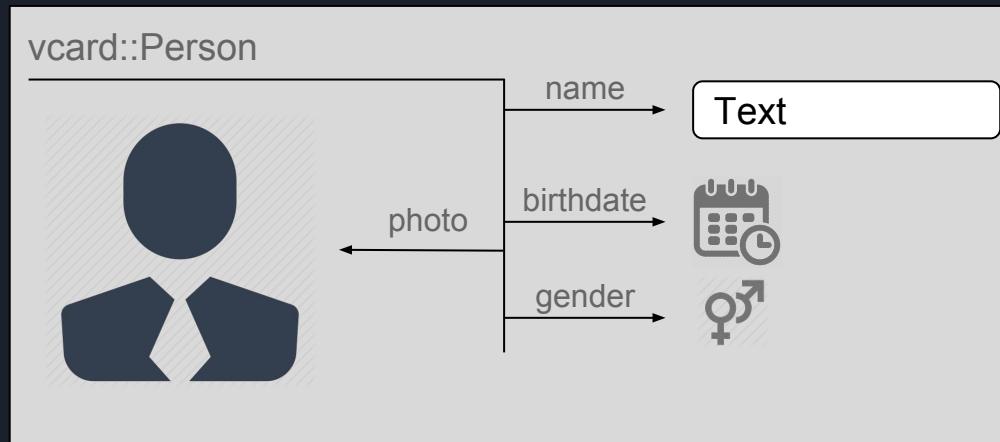
Completo: todo el dominio de conocimiento

Disjunto: nada puede ser dos cosas



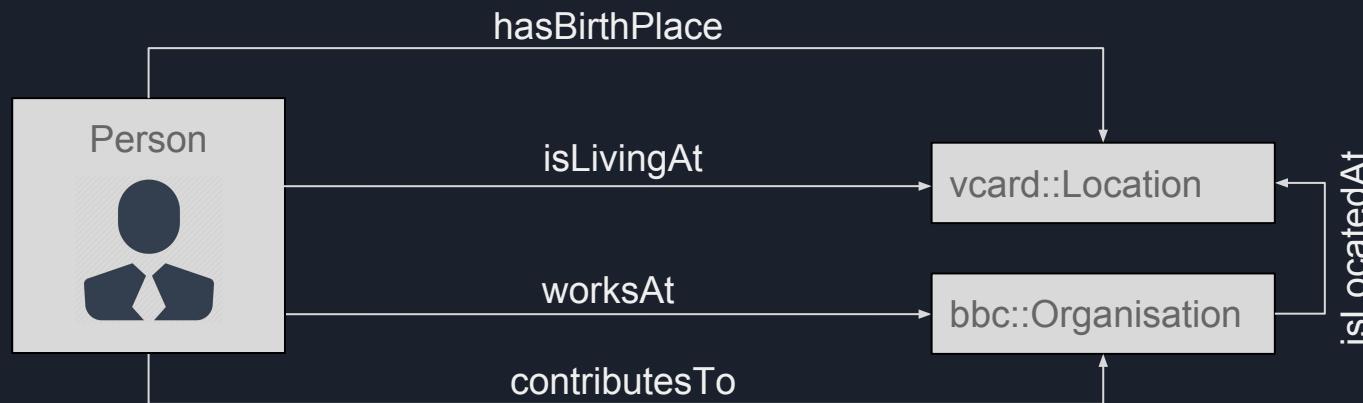
# Ontología. Propiedades - Atributos

- Definen características de un único concepto
- Información concreta



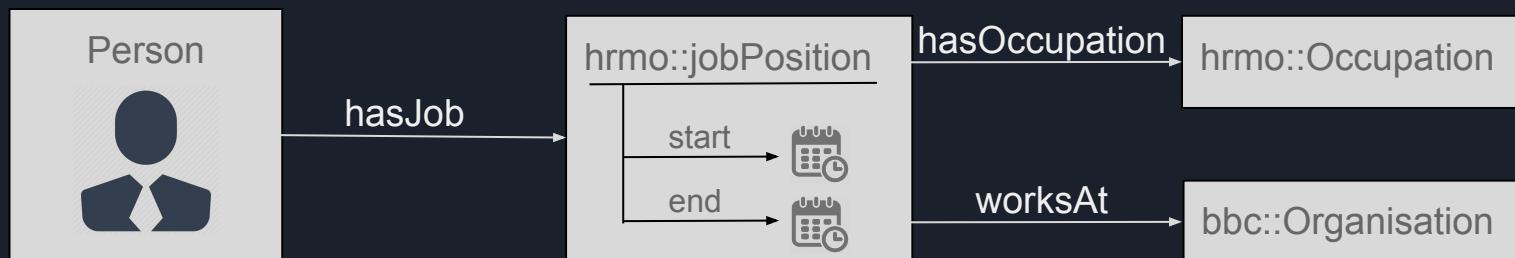
# Ontología. Relaciones

Pueden existir varias relaciones entre dos mismos conceptos, cada una de ellas con significado diferente



# Ontología. Conceptos complejos

Codifican relaciones complejas entre más de dos conceptos:



# LIKE A BOSS

Es tu ontología y defines:

- lo que quieres
- como quieres





**“Happiness is only real when shared”**

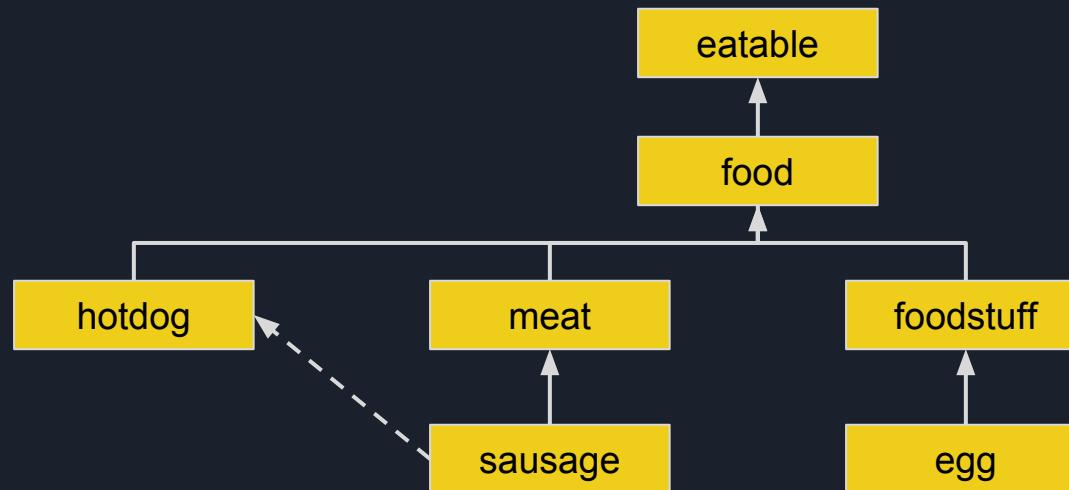
# WordNet.

## La ontología del lenguaje



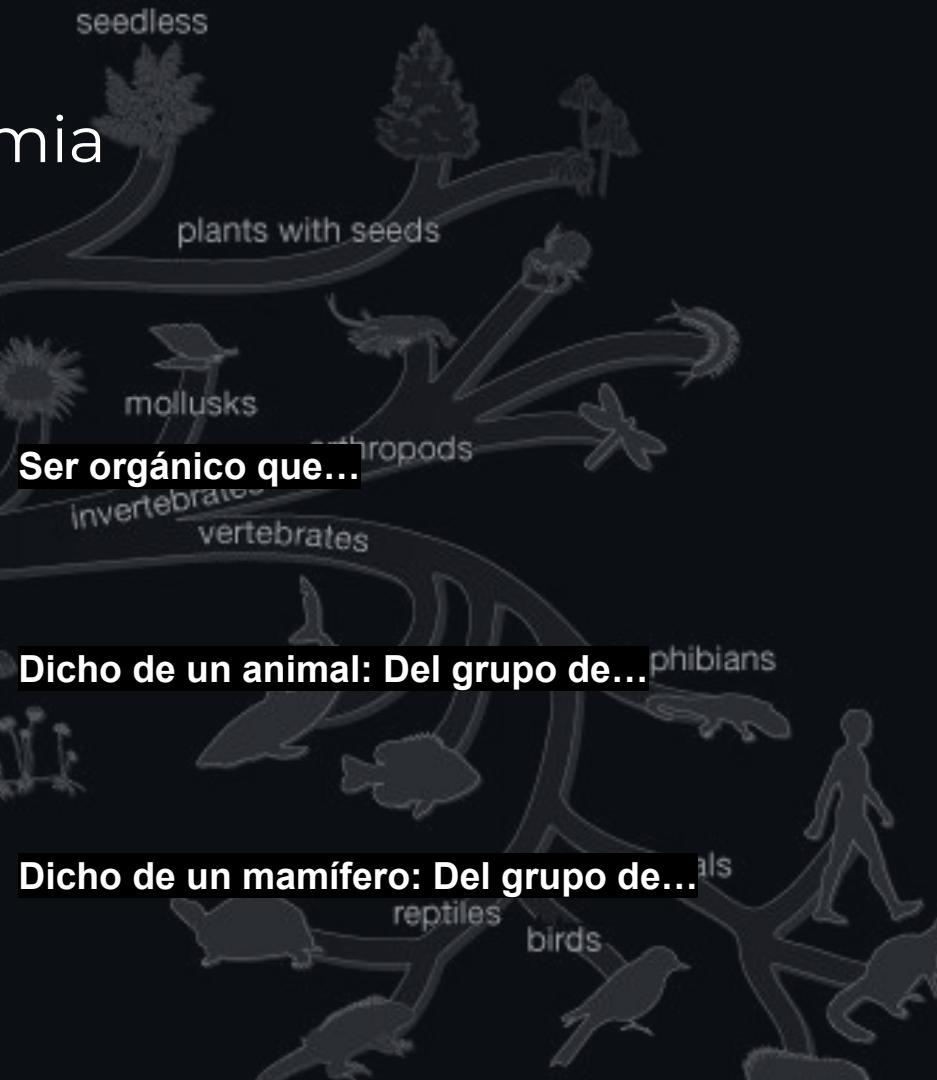
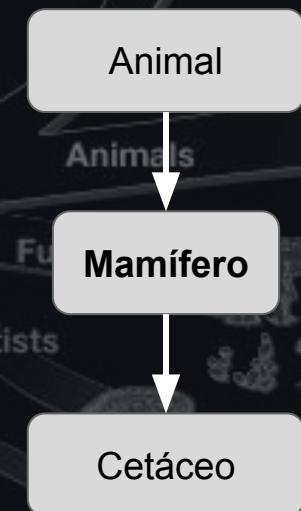
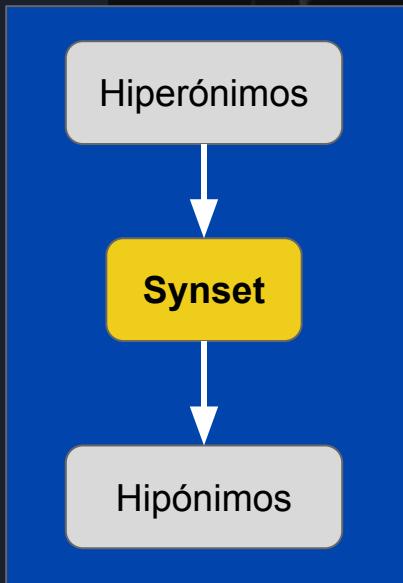
# WordNet. La ontología del lenguaje

Red de **conceptos** (synsets) con relaciones entre ellos: sinonimia, hiperonimia, hiponimia,...



# Hiperónimia - Hipónimia

Concreción en significado



# Holónimia - Meronimia

El todo y la parte

Holónimos

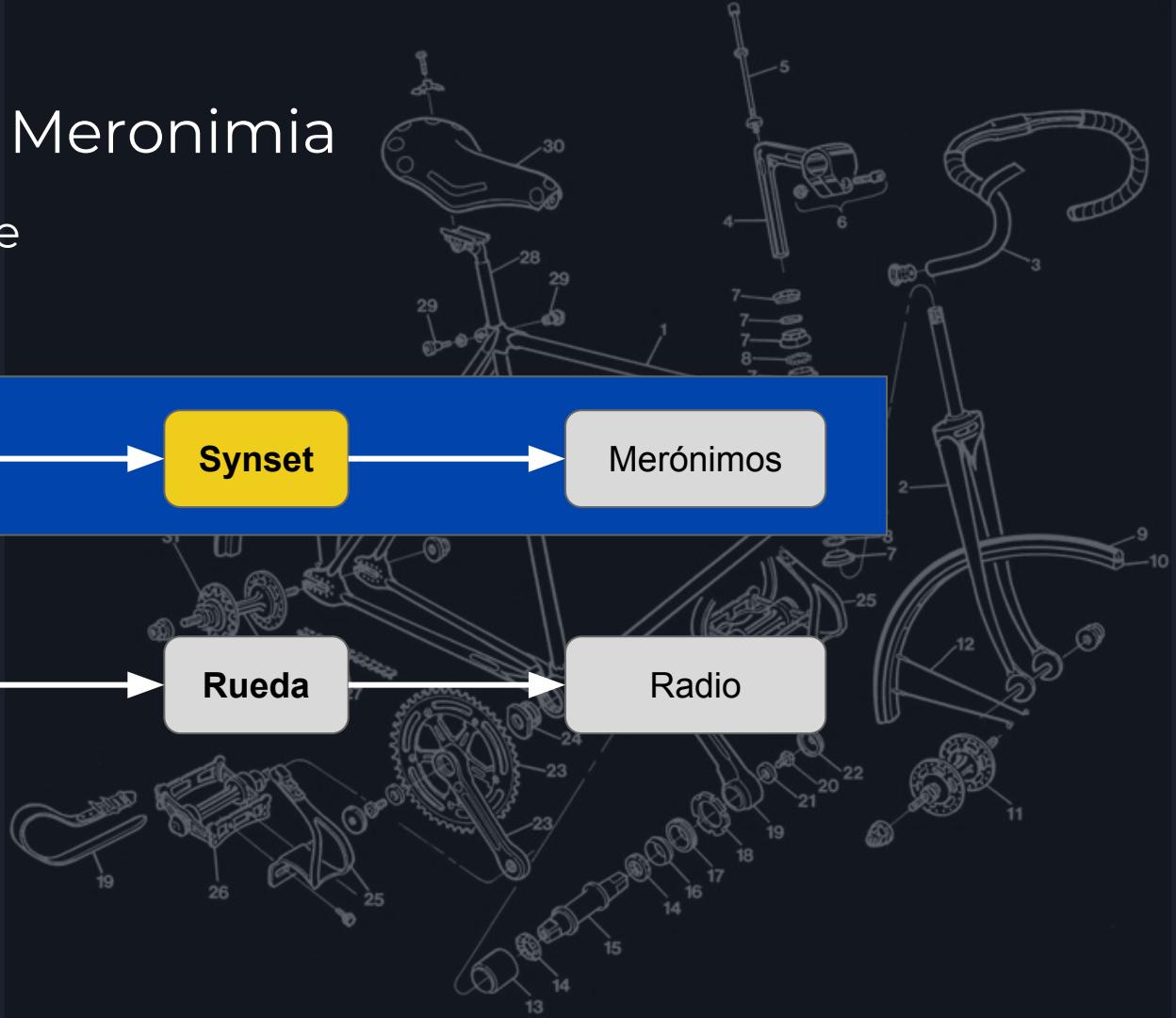
Synset

Merónimos

Bicicleta

Rueda

Radio



# 1 palabra/lema ⇒ 3 conceptos

## WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

### Noun

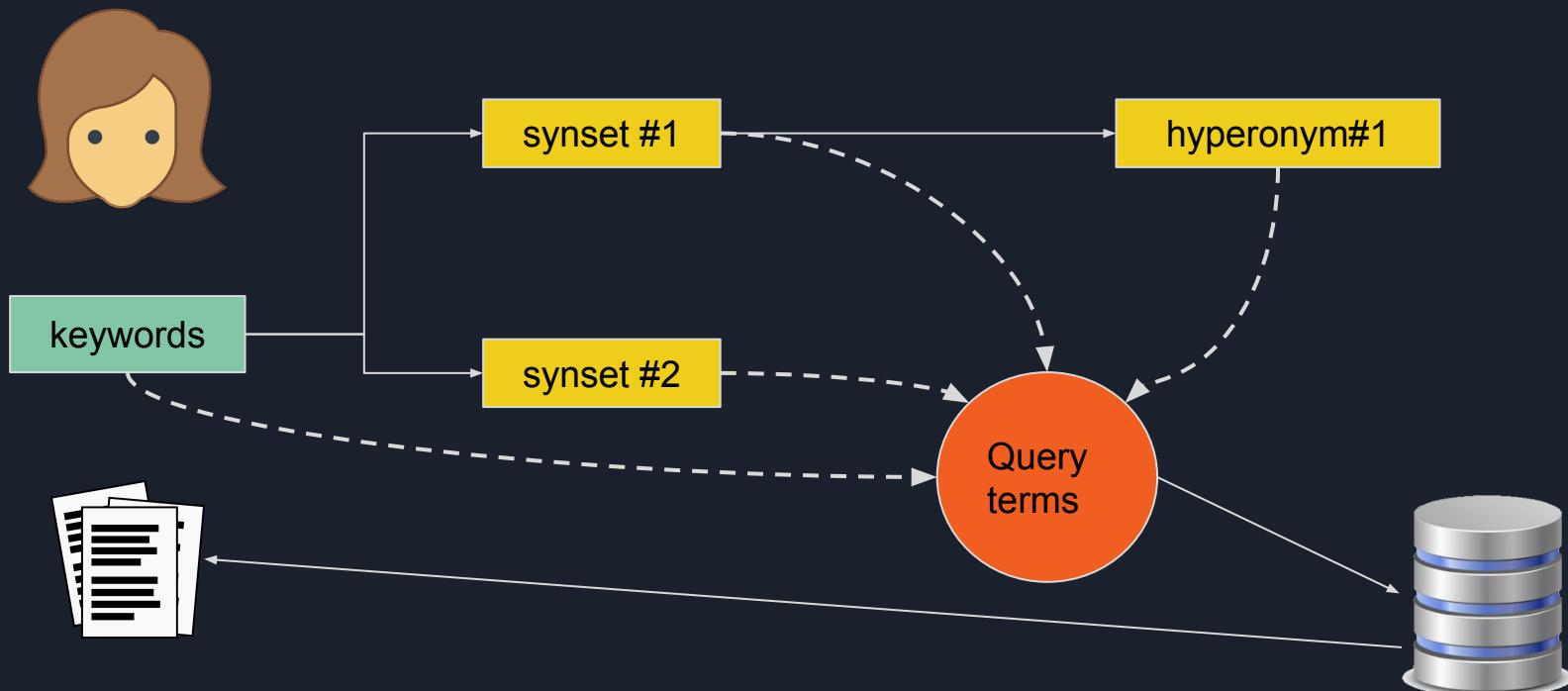
- {10207329} <noun.person>[18] S: (n) **hotdog#1 (hotdog%1:18:00::)**, hot dog#1 (hot\_dog%1:18:00::) (someone who performs dangerous stunts to attract attention to himself)
- {07713282} <noun.food>[13] S: (n) **hotdog#2 (hotdog%1:13:02::)**, hot dog#2 (hot\_dog%1:13:02::), red hot#1 (red\_hot%1:13:00::) (a frankfurter served hot on a bun)
- {07692347} <noun.food>[13] S: (n) **frank#2 (frank%1:13:00::)**, frankfurter#1 (frankfurter%1:13:00::), **hotdog#3 (hotdog%1:13:01::)**, hot dog#3 (hot\_dog%1:13:01::), dog#5 (dog%1:13:01::), wiener#2 (wiener%1:13:00::), wienerwurst#1 (wienerwurst%1:13:00::), weenie#1 (weenie%1:13:00::) (a smooth-textured sausage of minced beef or pork usually smoked; often served on a bread roll)

# Hiperonimia del concepto “hotdog#1”

- {10207329} <noun.person>[18] S: (n) **hotdog#1 (hotdog%1:18:00::)**, hot dog#1 (hot\_dog%1:18:00::) (someone who performs dangerous stunts to attract attention to himself)
  - direct hypernym / inherited hypernym / sister term
    - {10090518} <noun.person>[18] S: (n) exhibitionist#2 (exhibitionist%1:18:00::), show-off#1 (show-off%1:18:00::) (someone who deliberately behaves in such a way as to attract attention)
      - {10067267} <noun.person>[18] S: (n) egotist#1 (egotist%1:18:00::), egoist#1 (egoist%1:18:01::), swellhead#1 (swellhead%1:18:00::) (a conceited and self-centered person)
        - {09654651} <noun.person>[18] S: (n) unpleasant person#1 (unpleasant\_person%1:18:00::), disagreeable person#1 (disagreeable\_person%1:18:00::) (a person who is not pleasant or agreeable)
          - {09654317} <noun.person>[18] S: (n) unwelcome person#1 (unwelcome\_person%1:18:00::), persona non grata#2 (persona\_non\_grata%1:18:00::) (a person who for some reason is not wanted or welcome)
            - {00007846} <noun.Tops>[03] S: (n) person#1 (person%1:03:00::), individual#1 (individual%1:03:00::), someone#1 (someone%1:03:00::), somebody#1 (somebody%1:03:00::)

# Recuperación de información

## Expansión de consultas



# Recuperación de información

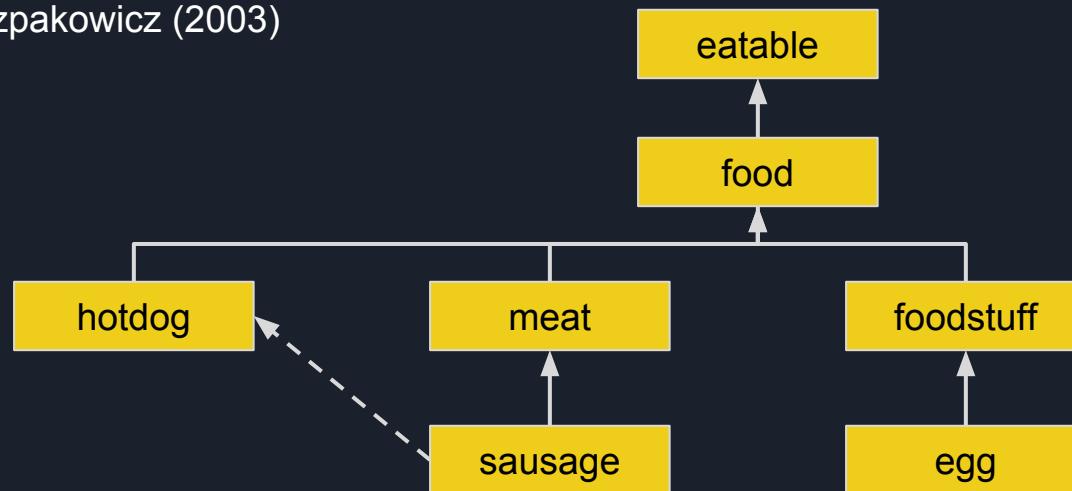
## Expansión de consultas



# Distancia entre conceptos

Contar conexiones en el grafo:

- Rada et al. (1989)
- Jarmasz y Szpakowicz (2003)





# Distancia entre conceptos

Intuitivamente no todas las conexiones deberían tener el mismo peso, pero no podemos etiquetarlas a mano. Algunas ideas:

- **Densidad**: a mayor densidad, menor distancia semántica entre los nodos de la zona.
- **Profundidad**: cuanto más se desciende, más sutiles son las diferencias, así que menor distancia.
- **Fuerza de cada conexión**: la relación de un padre con cada uno de sus hijos puede tener distinto peso (contenido de información).

# WordNet in a nutshell

# El filólogo estructurado



# PLN. ¿Cuál es el problema fundamental?

Tokenizar

Pos-tagging

Resolver correferencias

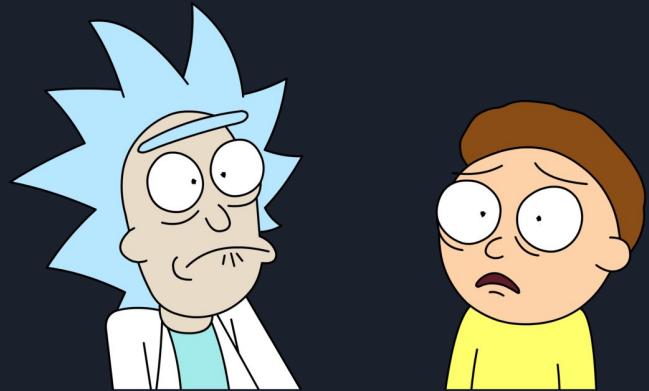
Reconocimiento de entidades

Codificación en una estructura



Ontologías

Knowledge databases



# PLN. ¿Cuál es el problema fundamental?

Tokenizar

Pos-tagging

Resolver correferencias

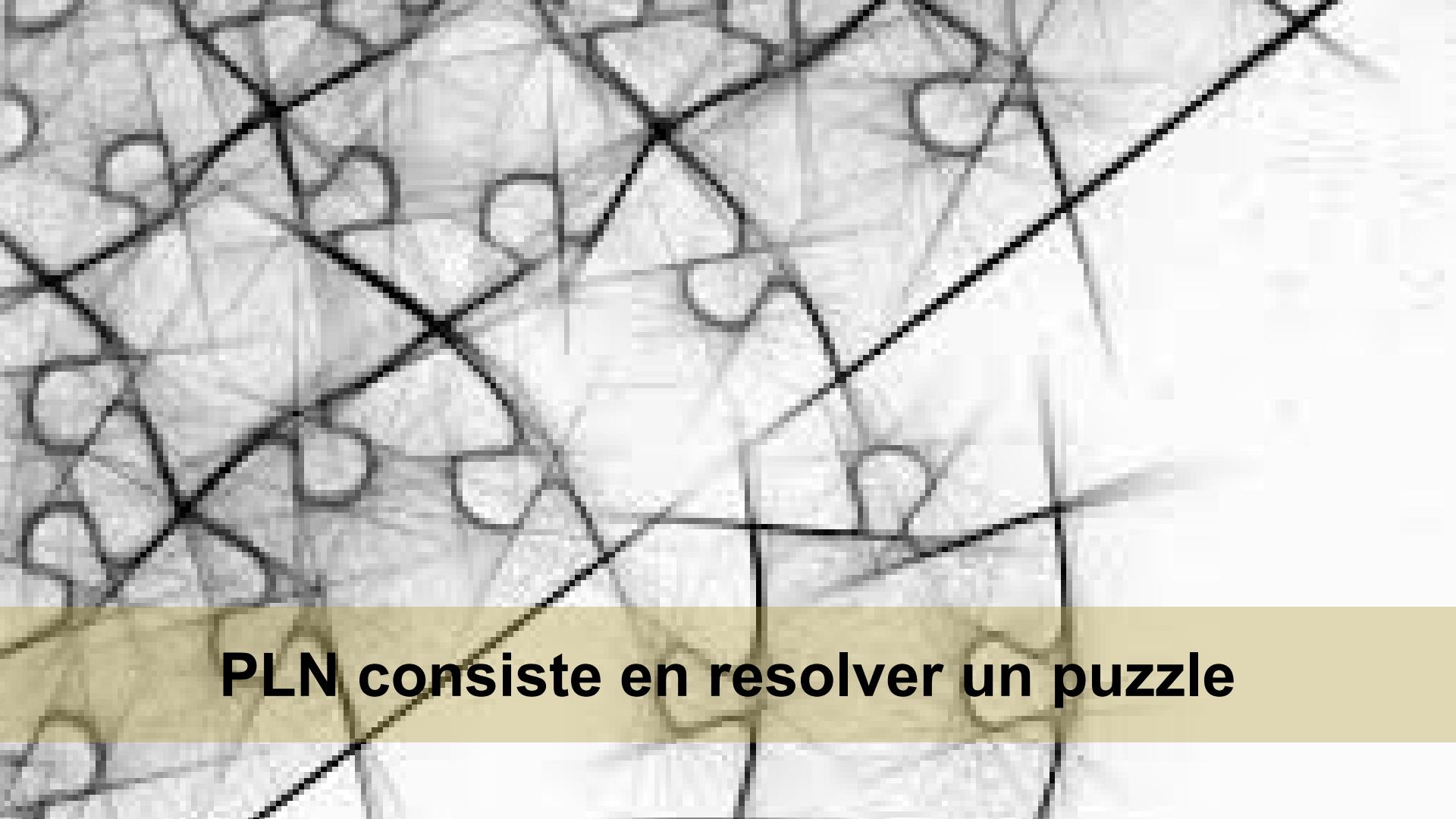
Reconocimiento de entidades

Codificación en una estructura

ontologías

Knowledge databases

**DESOMBICUAR**



**PLN consiste en resolver un puzzle**

# Resolver ambigüedades

Cristina Vela trabaja en la Universidad de Valladolid

en

la

----

trabajar

Tengo 18 significados posibles

U. Valladolid

DBpedia me dice que es una [Thing >> Agent >> Organization]

Cristina Vela

⇒ tiene que ser una [Agent >> Living thing >> Animal >> Person]



**¡No hace falta resolver todo el puzzle!**

# El filólogo estructurado

Recoger todo el conocimiento de un dominio

Resumirlo en conceptos y relaciones

Codificarlo para una máquina (base de datos)

Explotar su utilización



Imagen: Picasso. Retrato de Wilhelm Uhde. 1910



# ¿Por qué un filólogo?

**Trabajar:**

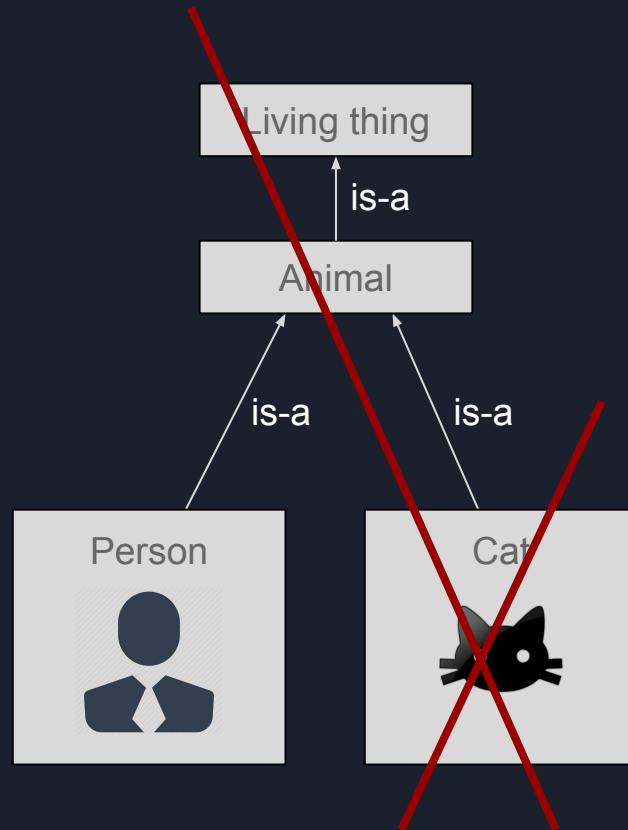
2. tener una ocupación remunerada en una empresa, una institución, etc.. (DLE)

# ¿Por qué un filólogo?

Trabajar:

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

- Agente: Animal >> Person

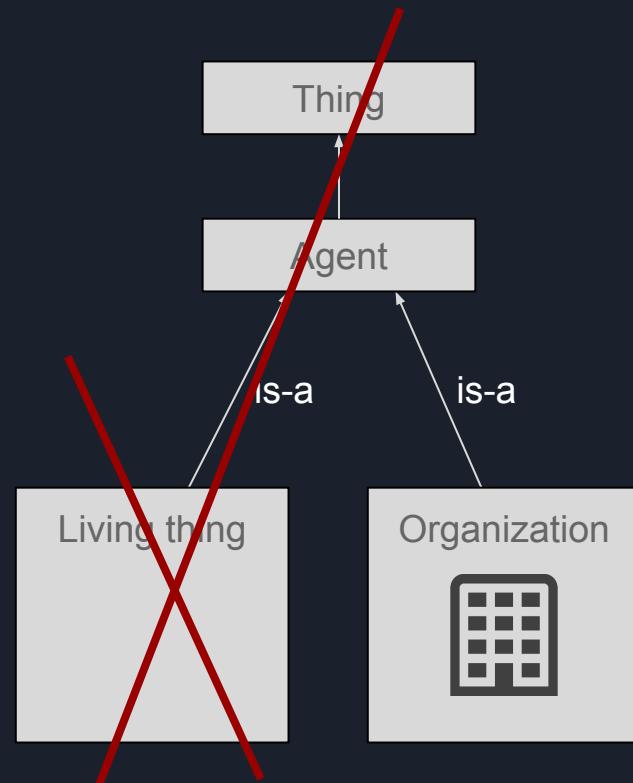


# ¿Por qué un filólogo?

Trabajar:

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

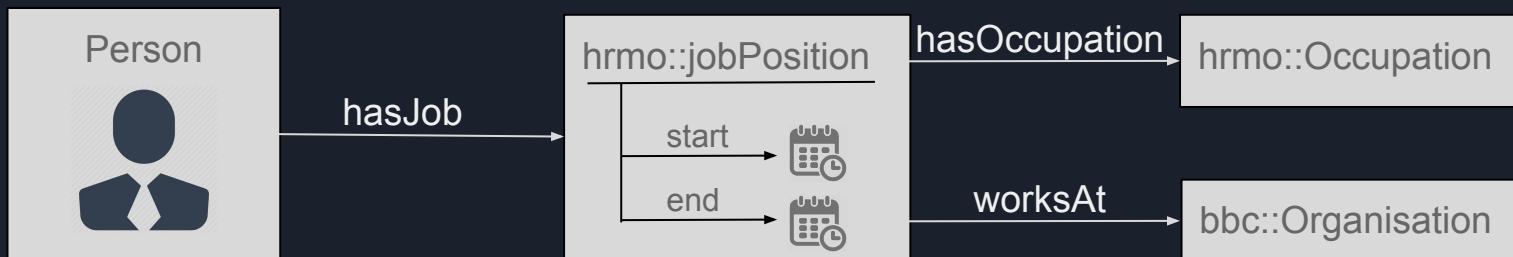
- Agente: Animal >> Person
- C1: Thing >> Agent >> Organization
- C2: Thing >> Work (occupation)



# ¿Por qué un filólogo?

## Trabajar

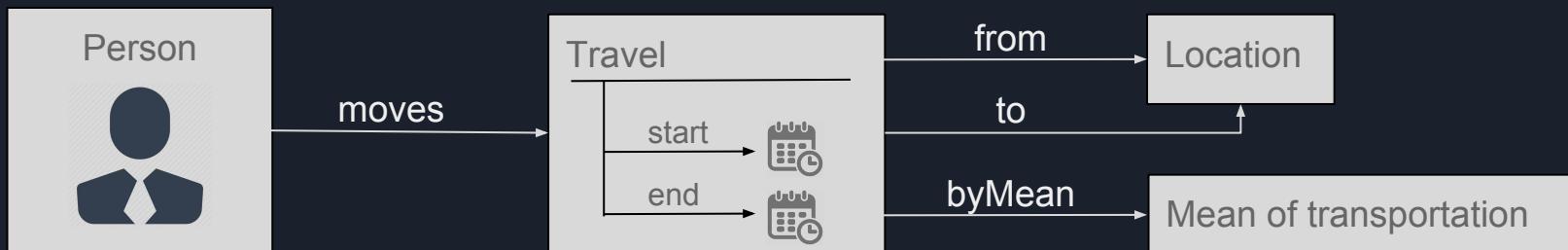
2. tener una ocupación remunerada en una empresa, una institución, etc.. (DLE)



# ¿Por qué un filólogo?

## Viajar

1. Trasladarse de un lugar a otro, generalmente distante, por cualquier medio de locomoción. (DLE)



# ¿Por qué un filólogo?

## Entregar

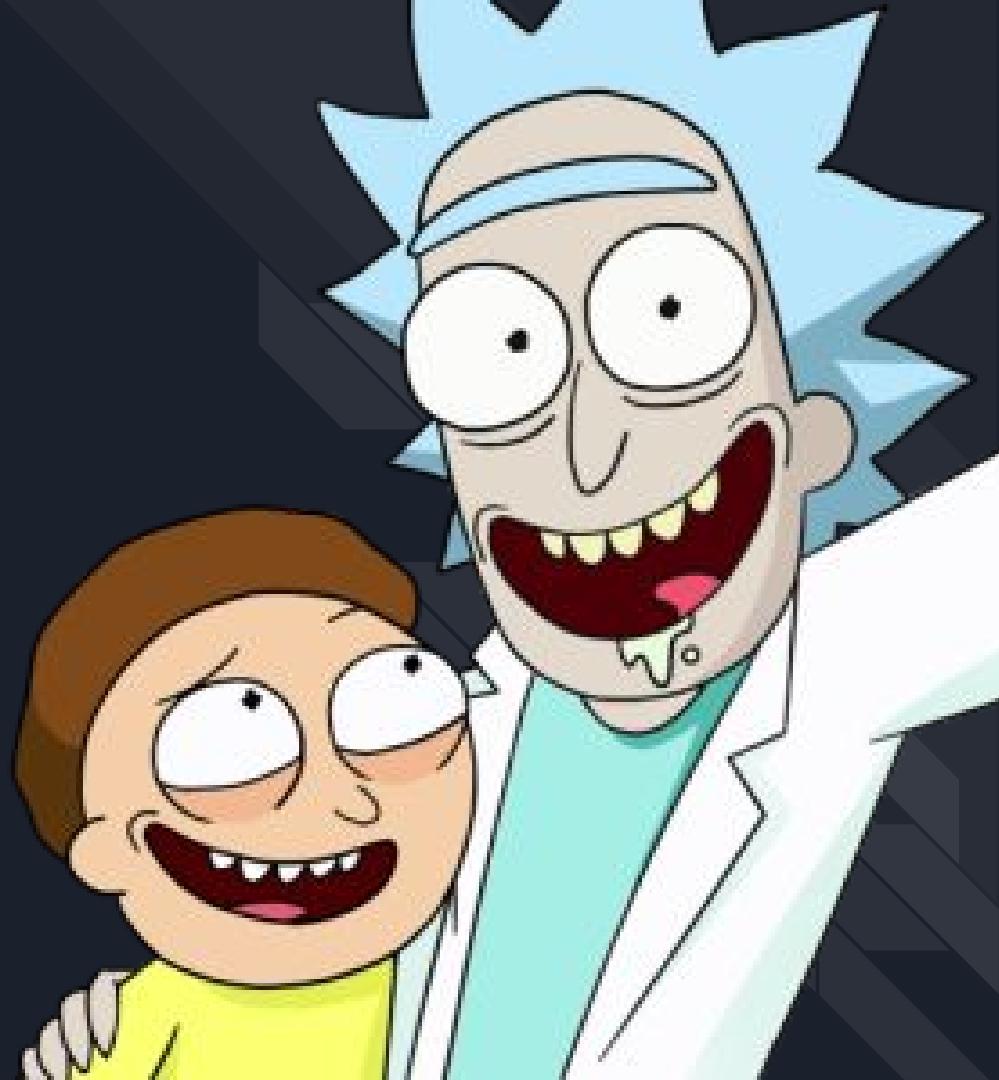
1. tr. Dar algo a alguien, o hacer que pase a tenerlo. (DLE)





**Una base de datos con la mente del filólogo**

PLN con ontologías



# PLN con ontologías



WordNet.- Desambiguar conceptos

Ontología de dominio (la mía)...

... mapeada con ontologías externas

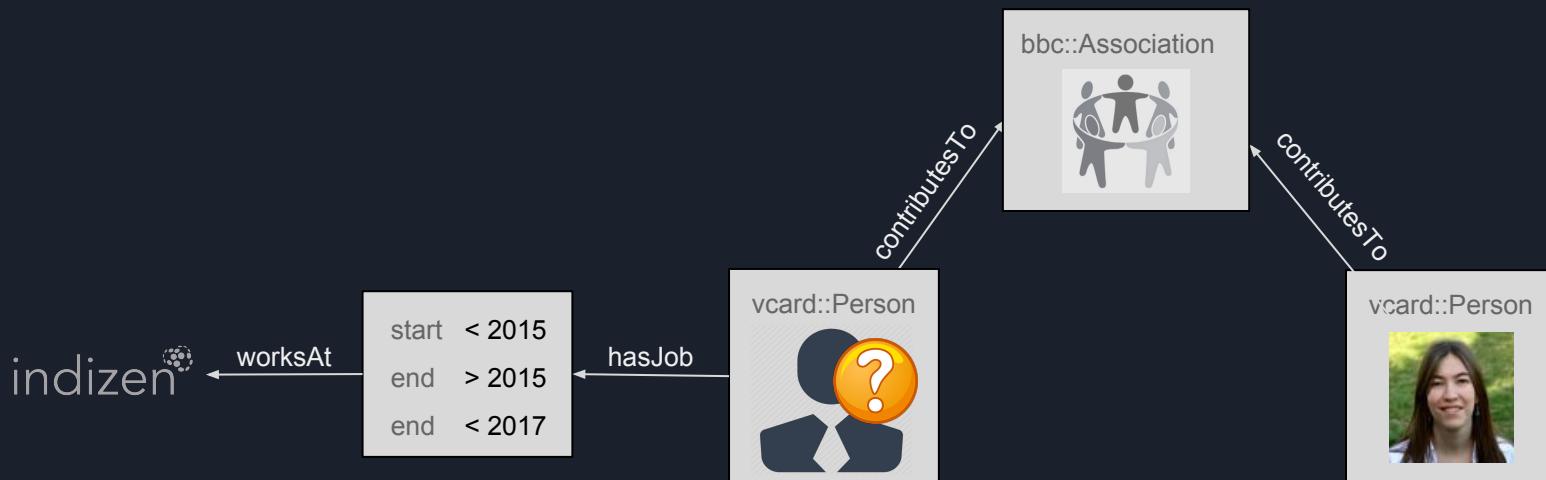
# PLN con ontologías

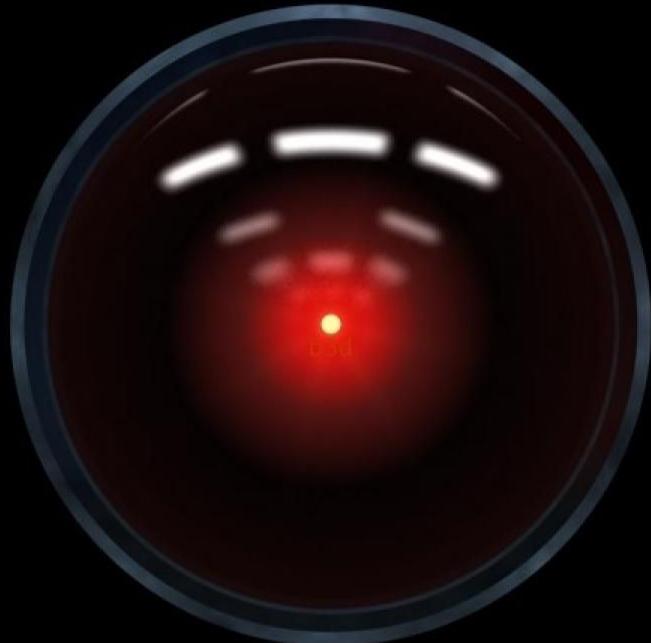
La ontología como interlingua  $\Rightarrow$  generador de lenguaje



# PLN con ontologías

La ontología como motor de inferencias





La web semántica como base de datos global

# Muchas gracias

## ¿Preguntas?



Javier G. Sogo  
[@jgsogo](https://twitter.com/jgsogo)



Lingwars  
[@lingwars](https://twitter.com/lingwars)

