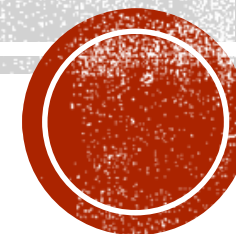


NLP

&

ONTOLOGIES

La lingüística en ayuda de la inteligencia artificial



Lorena Giusio

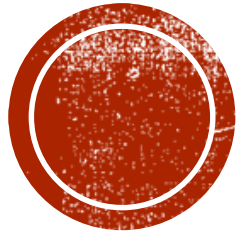
– NLU – Computational Linguist

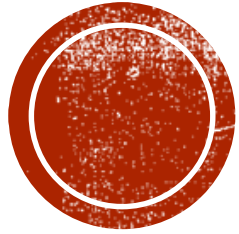


Linkedin: www.linkedin.com/in/lorenagiusio

Twitter: acetoarcobaleno

¿QUÉ ES EL NATURAL LANGUAGE PROCESSING?





- Traducción automática:
 - Search engine (Google Search, Lucene)
 - Text database (MongoDB, CouchDB)
- Clasificación de texto:
 - Resumen automático, palabras clave, categorías, entidades, fechas, direcciones, emails. Scraping
- Recuperación de información:
 - Question & Answering
 - Chatbots
- Extracción de información
- Entendimiento / Generación de lenguaje natural
- Corrección automática
- Compiladores / Intérpretes
-

APLICACIONES

HISTORIA

**“Una computadora
puede ser llamada
inteligente
si logra engañar
a una persona
haciéndole creer que es un
humano”**

- 50's – Alan Turing “Computing Machinery and Intelligence” – “Turing Test”
- '60 – Noam Chomsky, “Syntactic Structures”: sistemas basados en reglas. Traducción automática.
- 70's – Ontologías
- 80's – Sistemas basados en Aprendizaje Automático
- 90's – Mayor poder de cómputo.

Teorías Lingüísticas de Chomsky pierden fuerza. Modelos estadísticos.

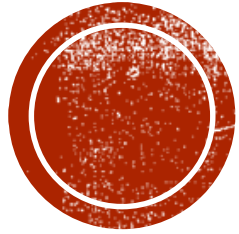
Alan Mathison Turing
1912–1954

HOY: CHAT BOT



El Chatbot es un **software de mensajería** que interactúa con usuarios a través del chat, ya sea de Facebook, sitio web o incluso a través de SMS, **simulando ser un operador o una persona real**. Estos sistemas de inteligencia artificial son excelentes para optimizar la experiencia del usuario, gestionar pedidos y resolver sus necesidades.





- Segmentación de oraciones
 - Un “.” puede ser decimal, suspensivo, abreviatura, etc.
- Segmentación de palabras (Tokenización)
 - En algunos lenguajes no es simple
- Análisis lexicográfico (segmentación de morfemas)
 - Stemming / Lemmatizing: Obtener forma canónica, lexema, raíz
 - Ej: run, runs, ran, running -> run
- Etiquetado gramatical (POS Tagging)
 - Asignar una etiqueta grammatical a una palabra
- Reconocimiento de entidades (NER)
- Análisis sintáctico (Parsing)
- Análisis semántico

TAREAS EN NLP

HERRAMIENTAS PARA ANALIZAR NLP

- Python -> muchas funciones para procesamiento de texto – open source
- Perl -> muchas funciones para procesamiento de texto – open source
- R -> Modelos estadístico(Machine Learning - Hidden Markov Model - Support Vector Machine (SVM) - Latent Semantic Analysis)
- NLTK
- **Protegé -> ontología (WordNet)**
- Expresiones Regulares
- Gramáticas
- Parser
- Clasificadores





¿QUÉ ES UNA ONTOLOGÍA?

El término **ONTOLOGIA** viene del campo de la filosofía, y se define como la rama de la filosofía que se ocupa de la naturaleza y **organización de la realidad**, es decir de **lo que "existe"**.

En el campo de la **Inteligencia Artificial** "lo que existe es **aquello que puede ser representado**".

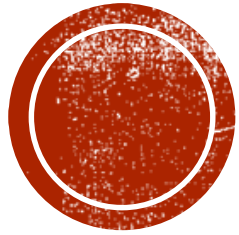
Definiciones:

Una ontología define el vocabulario de un área mediante un conjunto de términos básicos y relaciones entre dichos términos, así como las reglas que combinan términos y relaciones que amplían las definiciones dadas en el vocabulario.

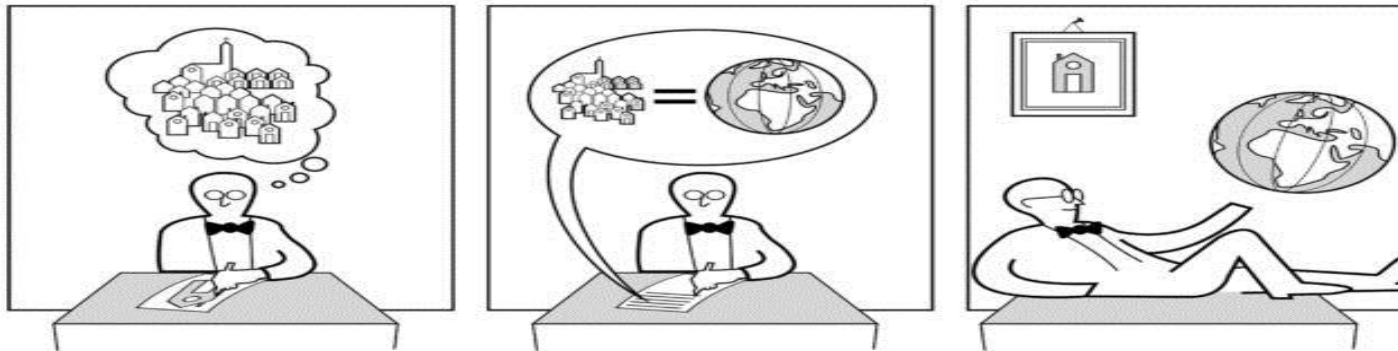
Una ontología cataloga las variables requeridas para algún conjunto de computación y establece las relaciones entre ellos.



¿A QUÉ/QUIÉN SIRVE UNA ONTOLOGÍA?



En los campos de la **inteligencia artificial, la Web Semántica, ingeniería de sistemas, ingeniería de software, informática biomédica, bibliotecología y arquitectura de la información** se crean ontologías para limitar la complejidad y para organizar la información. La ontología puede entonces ser aplicada para resolver problemas.



- Lo que muchas ontologías tienen en común, tanto en Informática como en Filosofía, es la **representación de entidades, ideas y eventos**, junto con sus propiedades y relaciones, de acuerdo con su sistema de categorización.
- La diferencia entre los dos es en la manera en que se enfocan. Los informáticos están más preocupados por un **establecimiento fijo y vocabularios controlados**, mientras que los filósofos están más preocupados por los principios, es decir si existen cosas tales como una esencia fija o si las entidades deben primar ontológicamente sobre los procesos.



¿DE QUÉ SE COMPONE UNA ONTOLOGÍA?

Aunque los autores pueden establecer varios criterios para definir una ontología, de manera general estas se componen de:

- **Dominio**, es el área de estudio que se desea explorar.

Por ejemplo, una ontología del dominio *Pizza* podría tener como objetivo definir todo lo que implica una pizza. El dominio puede ser tan vasto o tan cerrado como se desee y depende de la cantidad de información que se tenga disponible.

- **Concepto o clase**, es una idea que concibe o forma el conocimiento. Un concepto puede ser una definición proveniente de un grupo experto, un concepto del diccionario o un elemento de una base de datos.

Por ejemplo, el dominio *Pizza* está formado por conceptos como *Masa*, *Cobertura*, *Ingredientes*, etc.

- **Relación**, es la conexión existente entre conceptos.

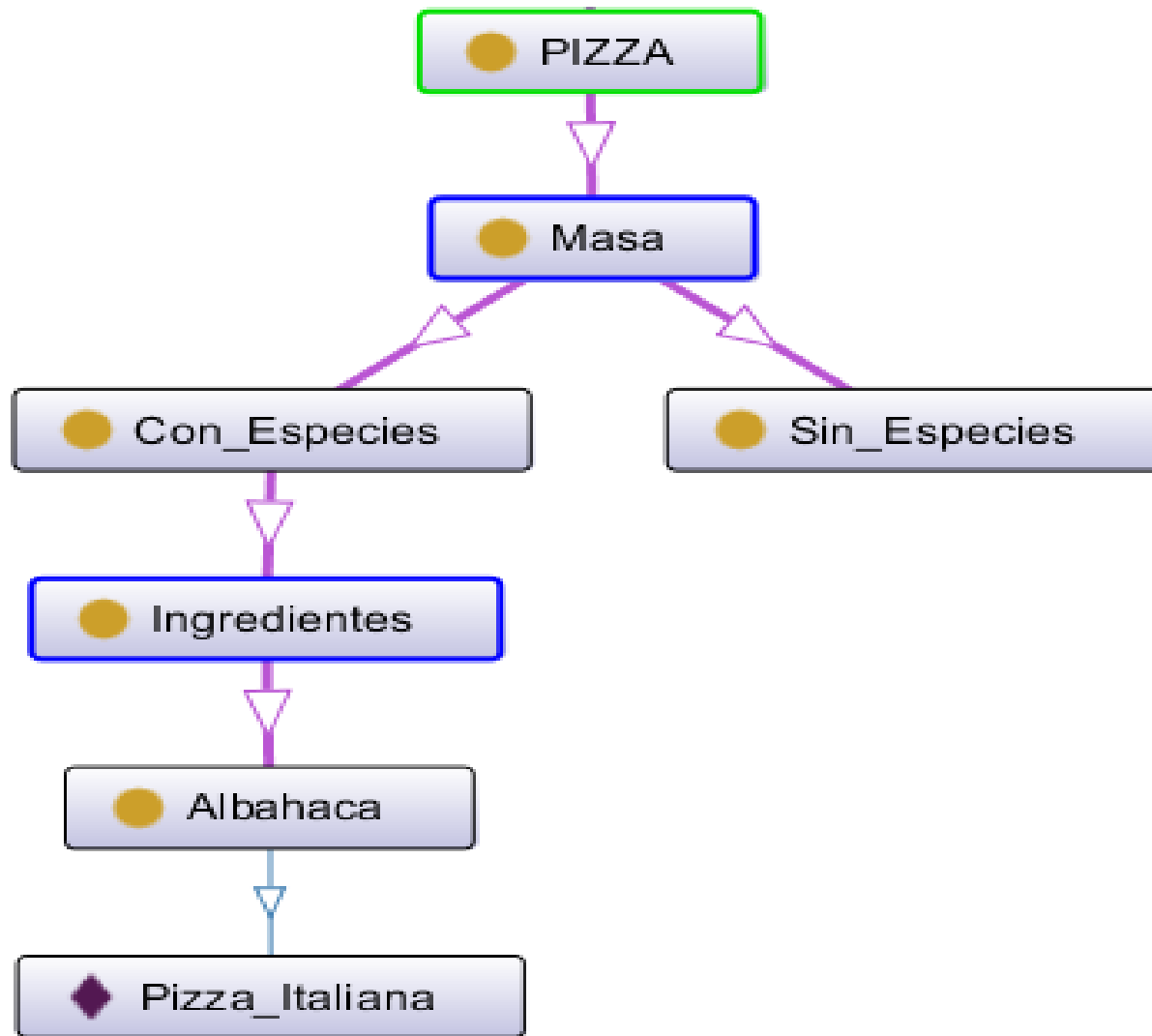
Por ejemplo, una *Pizza* ***está hecha de*** *Masa* y ***tiene una*** *Cobertura*, las frases remarcadas son la relación entre el dominio *Pizza*, el concepto *Masa* y el concepto *Cubierto*.

- **Instancias**, son los objetos que pueden definirse a través de los conceptos de una ontología.

Por ejemplo, *Pizza Italiana* es una instancia del dominio *Pizza* a través de conceptos específicos de las clases *Masa* e *Ingredientes*.



Pizza!!!!



El dominio o área de estudio de PIZZA, el cual tiene como primer concepto Masa (es decir, la relación es que una Pizza está hecha de Masa). Puede haber dos tipos de masa (dos nuevos conceptos que son hijos de Masa), a su vez la Masa-Con_Especies puede estar hecha de Ingredientes como la Albahaca. **Finalmente, la instancia Pizza_Italiana puede definirse como "Una PIZZA hecha de Masa-Con_Especies con Ingredientes como la Albahaca".**

Es decir, **todos los conceptos introducidos nos permitieron reconstruir conocimiento real.**



HERRAMIENTAS PARA CREAR ONTOLOGÍAS

Entendiendo los conceptos básicos de ontologías es posible desarrollarlas en muchas plataformas y formatos, por ejemplo con dibujos, diagramas o lenguajes de marcas o etiquetas (como XML).

Sin embargo, la Universidad de Standford ha desarrollado una herramienta llamada Protégé, un editor de ontologías gratuito y de código abierto.



http://www.co-ode.org/ontologies/pizza/pizza.owl - [http://www.co-ode.org/ontologies/pizza/2007/02/12/pizza.owl]

File Edit Reasoner Tools Refactor Tabs View Window Help

http://www.co-ode.org/ontologies/pizza/pizza.owl

Active Ontology Entities Classes Object Properties Data Properties Individuals OWL Viz DL Query

Asserted Class Hierarchy: Pizza

- Thing
 - DomainConcept
 - Country
 - Food
 - IceCream
 - Pizza**
 - CheeseyPizza
 - InterestingPizza
 - MeatyPizza
 - NamedPizza
 - NonVegetarianPizza
 - RealItalianPizza
 - SpicyPizza
 - SpicyPizzaEquivalent
 - ThinAndCrispyPizza
 - VegetarianPizza
 - VegetarianPizzaEquivalent
 - VegetarianPizzaEquivalent1
 - PizzaBase
 - PizzaTopping
 - ValuePartition

OWL Viz: Pizza

Show class Show children Show parents Show all classes Hide class Hide children Hide classes

Asserted model Inferred model

```
graph TD
    Thing --> DomainConcept
    DomainConcept --> Country
    Country --> Food
    Food --> IceCream
    Food --> Pizza
    Pizza --> CheeseyPizza
    Pizza --> InterestingPizza
    Pizza --> MeatyPizza
    Pizza --> NamedPizza
    Pizza --> NonVegetarianPizza
    Pizza --> RealItalianPizza
    Pizza --> SpicyPizza
    Pizza --> SpicyPizzaEquivalent
    Pizza --> ThinAndCrispyPizza
    Pizza --> VegetarianPizza
    Pizza --> VegetarianPizzaEquivalent
    Pizza --> VegetarianPizzaEquivalent1
    Pizza --> PizzaBase
    Pizza --> PizzaTopping
```



ONTOLOGÍAS DISPONIBLES

- **BabelNet:** es una red semántica multilingüe y una ontología lexicalizada. BabelNet ha sido creada automáticamente integrando la mayor enciclopedia web multilingüe, Wikipedia, a la base de datos léxica más reconocida del idioma inglés, WordNet.
- **Cyc:** a large Foundation Ontology para representación formal del universo de discurso .
- **Disease Ontology:** diseñado para facilitar el reconocimiento de enfermedades y condiciones asociadas a un código médico en particular.
- **Ontology for Biomedical Investigations:** un software de acceso abierto, ontología integrada para la descripción de investigaciones biológicas y clínicas.
- **Geopolitical ontology:** Es una ontología que describe información geopolítica creada por la FAO. La ontología geopolítica incluye nombres en diferentes idiomas (español, francés, español, árabe, chino, ruso e italiano); con los códigos de los sistemas (UN, ISO, FAOSTAT, AGROVOC, etc.); facilitando una relación entre territorios (fronteras, grupos de membresía, etc).



BIBLIOGRAFÍA

- Ontologías

Chandrasekaran B., Josephson J.R., Benjamins V.: Ontology of Tasks and Methods,. In Proceedings of KAW'98, Inn, Banff, Alberta, Canada, (1998).

Chandrasekaran B., Josephson J.R., Benjamins V.: What Are Ontologies, and Why Do We Need Them?. IEEE Intelligent Systems, v.14n.1, (1999) pp 20-26.

Gruninger, M. y Lee, J. (2002): Ontology Applications and Design. Communications of the ACM. 45(2), (2002) pp. 39-41

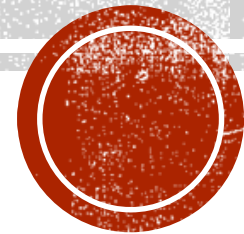
- NLP

Jurafsky, Dan, and James H. Martin. Speech and language processing. Vol. 3. London: Pearson, 2014.

Manning, Christopher D., and Hinrich Schütze. Foundations of statistical natural language processing. Vol. 999. Cambridge: MIT press, 1999.



GRACIAS!!!!



Lorena Giusio

– NLU – Computational Linguist

Linkedin: www.linkedin.com/in/lorenagiusio

Twitter: [acetoarcobaleno](https://twitter.com/acetoarcobaleno)