

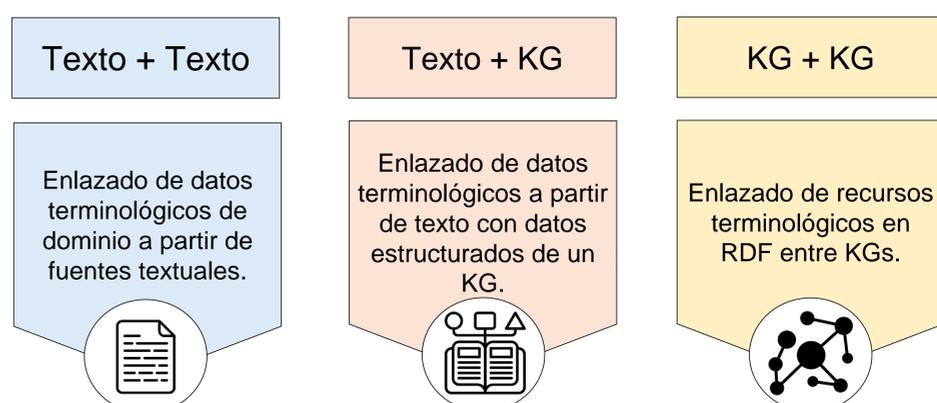
Desambiguación para el enlazado de datos de dominio mediante modelos de lenguaje

Lucía Palacios, lucia.palacios@upm.es

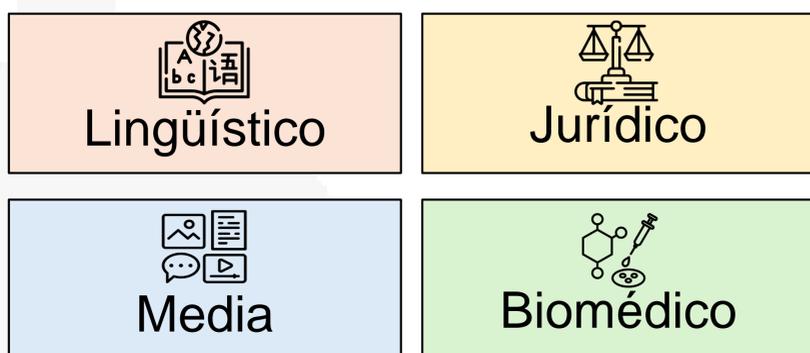
Objetivos

1. Ofrecer un servicio de desambiguación (más conocido como Word Sense Disambiguation-WSD) de recursos terminológicos de dominio en formato de datos enlazados.
2. Investigar técnicas de PLN que permitan entender y resolver la ambigüedad léxica del lenguaje, lo cual posibilita la generación de datos para la Web Semántica y la explotación de estos.
3. Desarrollo y evaluación de un modelo capaz de ofrecer un servicio de desambiguación para los términos en español de dominios de especialidad.

Casos de uso



Dominios



Recursos disponibles

Inventarios de sentidos:

- **BabelNet** [Navigli y Ponzetto, 2012]
- **Wikidata** [Vrandečić y Krötzsch, 2014]
- **ConceptNet** [Speer et al., 2018]
- **Wiktionary** [Blevins et al., 2021]

Datasets:

- **XL-WSD** [Pasini et al., 2021]

Gold Standard:

- **SemEval-13 Task 12** [Navigli, Jurgens, and Vannella, 2013]
- **SemEval-15 Task 13** [Moro and Navigli, 2015]

Enfoques

WSD basado en conocimiento

Algoritmos que explotan la información semántica de los grafos de conocimiento. Independientes de corpus de entrenamientos.



WSD Supervisado

Algoritmos basados en arquitecturas neuronales. Entienden la tarea de WSD como una tarea de clasificación.

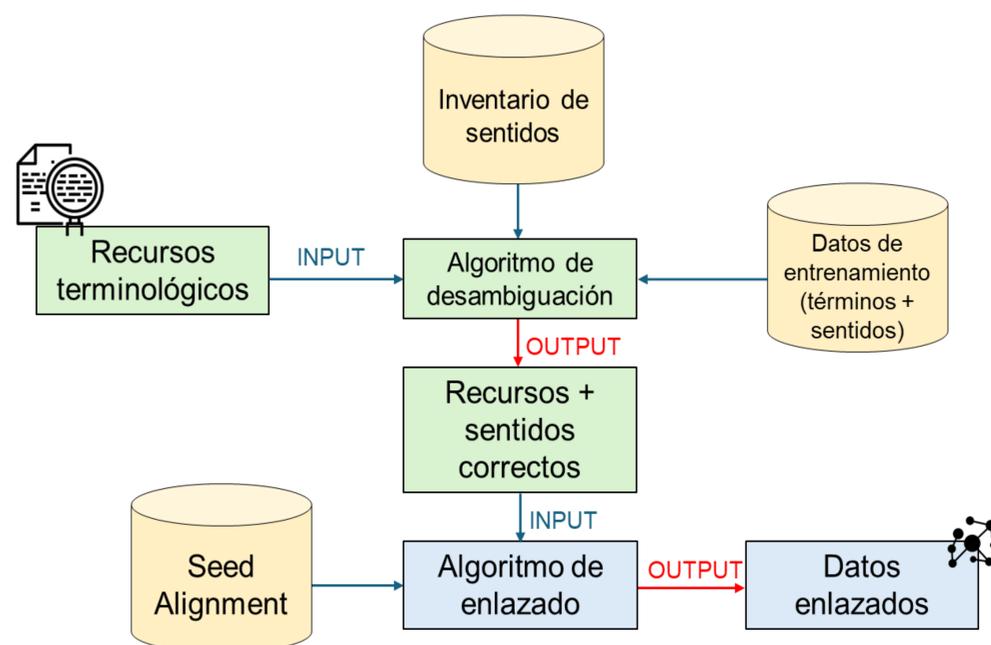


WSD Híbrido

Algoritmos que combinan los modelos de lenguaje con grafos de conocimiento.



Pipeline



Referencias

Bevilacqua, M., Pasini, T., Raganato, A., & Navigli, R. (2021). Recent Trends in Word Sense Disambiguation: A Survey. *International Joint Conference on Artificial Intelligence*.

Pasini, T. & Camacho-Collados, J. (2020). "A Short Survey on Sense-Annotated Corpora. In Proceedings of the Twelfth Language Resources and Evaluation Conference", pages 5759–5765, Marseille, France. European Language Resources Association.

Shen, W., Li, Y., Liu, Y., Han, J., Wang, J., and Yuan, X. (2021). "Entity Linking Meets Deep Learning: Techniques and Solutions," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 3, pp. 2556-2578.

Diagrams: Original artwork. Icons: Flaticon.com

Pasos futuros

1. WSD Cross-lingual y multimodal
2. WSD con LLMs
3. Resolución de problemas de granularidad y amplitud de inventarios de sentido: **Word Sense Extension**
4. Recursos de **desambiguación para el español específicos de dominio**
5. **Entity Linking sin supervisor**

