



Título: BIOMEDICAL ENTITIES RECOGNITION IN SPANISH COMBINING WORD EMBEDDINGS

Nombre: LÓPEZ ÚBEDA, PILAR

Universidad: Universidad de Jaén

Departamento: Informática

Fecha de lectura: 22/04/2021

Mención a doctor europeo: concedido

Programa de doctorado: Programa de Doctorado en Tecnologías de la Información y de la Comunicación por la Universidad de Jaén

Dirección:

- > **Director:** LUIS ALFONSO UREÑA LÓPEZ
- > **Codirector:** MARÍA TERESA MARTÍN VALDIVIA
- > **Codirector:** MANUEL CARLOS DÍAZ GALIANO

Tribunal:

- > **presidente:** RAFAEL MUÑOZ GUILLENA
- > **secretario:** PALOMA MARTÍNEZ FERNANDEZ
- > **vocal:** MANUEL MONTES GÓMEZ

Descriptores:

- > INTELIGENCIA ARTIFICIAL
- > INFORMÁTICA

El fichero de tesis ya ha sido incorporado al sistema

Localización: BIBLIOTECA UJA

Resumen: El reconocimiento de entidades con nombre (NER) es una tarea importante en el campo del Procesamiento del Lenguaje Natural que se utiliza para extraer conocimiento significativo de los documentos textuales. El objetivo de NER es identificar trozos de texto que se refieran a entidades específicas.

En esta tesis pretendemos abordar la tarea de NER en el dominio biomédico y en español. En este dominio las entidades pueden referirse a nombres de fármacos, síntomas y enfermedades y ofrecen un conocimiento valioso a los expertos sanitarios. Para ello, proponemos un modelo basado en redes neuronales y empleamos una combinación de word embeddings. Además, nosotros generamos unos nuevos embeddings específicos del dominio y del idioma para comprobar su eficacia. Finalmente, demostramos que la combinación de diferentes word embeddings como entrada a la red neuronal mejora los resultados del estado de la cuestión en los escenarios aplicados.



Named Entity Recognition (NER) is an important task in the field of Natural Language Processing that is used to extract meaningful knowledge from textual documents. The goal of NER is to identify text fragments that refer to specific entities.

In this thesis we aim to address the task of NER in the Spanish biomedical domain. In this domain entities can refer to drug, symptom and disease names and offer valuable knowledge to health experts. For this purpose, we propose a model based on neural networks and employ a combination of word embeddings. In addition, we generate new domain- and language-specific embeddings to test their effectiveness. Finally, we show that the combination of different word embeddings as input to the neural network improves the state-of-the-art results in the applied scenarios.