



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud


Guías de anotación

Plan de Impulso de las Tecnologías del Lenguaje

[M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima]¹ y [V. Briva-Iglesias]²


¹ Text Mining Unit, BSC

² D-REAL, Dublin City University

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 1/27

ÍNDICE

1	Introducción	2
2	Anotación	4
	Etiquetas de brat (Tabla 2)	4
	Reglas generales (Tabla 3)	7
	Reglas positivas (Tabla 4)	9
	Reglas negativas (Tabla 5)	14
	Casos especiales (Tabla 6)	16
3	Codificación	19

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 2/27

1. INTRODUCCIÓN

El propósito de esta guía es facilitar la detección, codificación y clasificación de profesiones/ocupaciones y situaciones laborales expresadas en lenguaje natural (castellano, catalán e inglés) y contenidas principalmente en un contexto de redes sociales («Social Media»). La detección y codificación se podrá usar posteriormente para determinar grupos ocupacionales de mayor riesgo de: 1) enfermedades transmisibles; 2) otras enfermedades agudas y crónicas; 3) trastornos de la salud mental; 4) conductas de riesgo, etc.

Para obtener un «Gold Standard» utilizaremos métodos de aprendizaje automático semisupervisado («semi-supervised machine learning»), en los cuales anotadores profesionales corregirán y codificarán la preanotación automática de aproximadamente 10 000 tuits bajo supervisión experta.


Se considerarán variantes de las lenguas mencionadas en los tuits. Por ejemplo, tendremos en cuenta que la expresión rioplatense *conductor radial* es equivalente a *locutor de radio* en castellano. Además, crearemos un diccionario trilingüe de profesiones que abarque el mayor número de ocupaciones tal y como se expresan en las redes sociales, añadiendo neologismos como por ejemplo, *gamer*, que se encuentran ausentes en las mayores clasificaciones.

Para la clasificación, codificación y «mapping» de resultados nos basaremos sobre todo en (por orden de prioridad):

1	ESCO	https://ec.europa.eu/esco/portal/occupation
2	SNOMED CT (contexto social/ocupación)	https://browser.ihtsdotools.org/?perspective=full&conceptId1=14679004&edition=MAIN/SNOMEDCT-ES/2020-04-30&release=&languages=es,en

Todas las ocupaciones en ESCO están mapeadas exactamente a un código ISCO-08. Organizaciones como el Instituto Nacional de Estadística utilizan la CNO-11, que se encuadra dentro del marco conceptual de la CIUO-08. El Idescat, con el asesoramiento del TERMCAT³,

³ <https://www.termcat.cat/en/node/559>

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 3/27

utiliza la CCO-2011, la traducción al catalán de la Clasificación nacional de ocupaciones 2011 (CNO-2011).

La persona anotadora y codificadora tendrá experiencia en anotación y/o traducción y tratamiento de textos, con conocimiento mínimo de 2 lenguas, una de ellas el inglés. Además, antes de anotar y codificar tendrá que familiarizarse con la clasificación ESCO de ocupaciones. Por último, el código de colores con el que se señalan los diferentes tipos de ejemplos en estas guías se muestra en la Tabla 1:

TABLA 1: CONVENCIONES EN LAS GUÍAS


1	Código cromático: anotación incorrecta
2	Código cromático por TAGs: Profesión ; Actividad ; Situación Laboral ; Profesión figurativa

2. ANOTACIÓN

Para la anotación utilizaremos la herramienta brat, en la que vamos a disponer de las siguientes categorías a la hora de anotar una ocupación:

TABLA 2. ETIQUETAS DE LA HERRAMIENTA BRAT

TAG	Incluye	Ejemplos
PROFESIÓN Ocupaciones que proporcionan a la persona un salario o modo de sustento	Profesiones convencionales; nuevas profesiones relacionadas con Internet; manteros; trabajadores sexuales; traficantes y otras ocupaciones ilegales. En el caso de ex- y co-, incluiremos los prefijos y consideraremos la mención como la misma profesión.	<i>periodista de Radio Amanecer</i> <i>Futura ingeniera biomédica</i> <i>co-director ejecutivo</i>
SITUACIÓN LABORAL	Jubilada; en paro; estudiante/doctorando/aprendiz/en formación; opositor; ERTE; autónomo; en baja maternal/paternal; esclavos; sin techo/mendigo; trabajador/a; profesional (sin apellido).	<i>Estudiante RRPP</i> 92.454 <i>ERTES</i> y 678.700 <i>parados</i> la <i>trabajadora</i> , <i>auxiliar administrativa</i> , de 41 años de edad

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 5/27

ACTIVIDAD Ocupaciones no remuneradas	Trabajadores del hogar; cuidador no remunerado; deportista o artista no profesionales; cargos comunitarios no remunerados; activista; voluntario; gurú; «gamer».	<i>Ama de casa venezolana</i> <i>Esquiador y fotógrafo</i> <i>presidente de la asociación de padres</i> <i>gamer impulsivo</i>
PROFESIÓN FIGURATIVA Uso metafórico de nombres de ocupaciones		<i>Pintor de techos de cine de verano</i> <i>Cowboy crepuscular</i> <i>Costalero de silencio</i>
_SUG_PROFESION	Las sugerencias de la etiqueta PROFESIÓN, aparecen como _SUG_PROFESION. El anotador cambia esta etiqueta por PROFESIÓN (o SITUACIÓN LABORAL, ACTIVIDAD o FIGURATIVA) en caso de que sea correcta.	
_REJ_PROFESION	Si es incorrecta, debe cambiarla a _REJ_PROFESION.	

A modo de ejemplo, abajo se muestran dos figuras: la primera muestra el resultado final de anotar un tuit y la segunda la interfaz de brat.

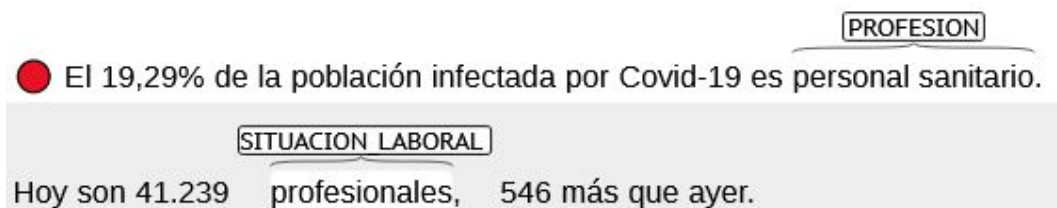



Figura 1: resultado final de anotación de un tuit.

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 6/27

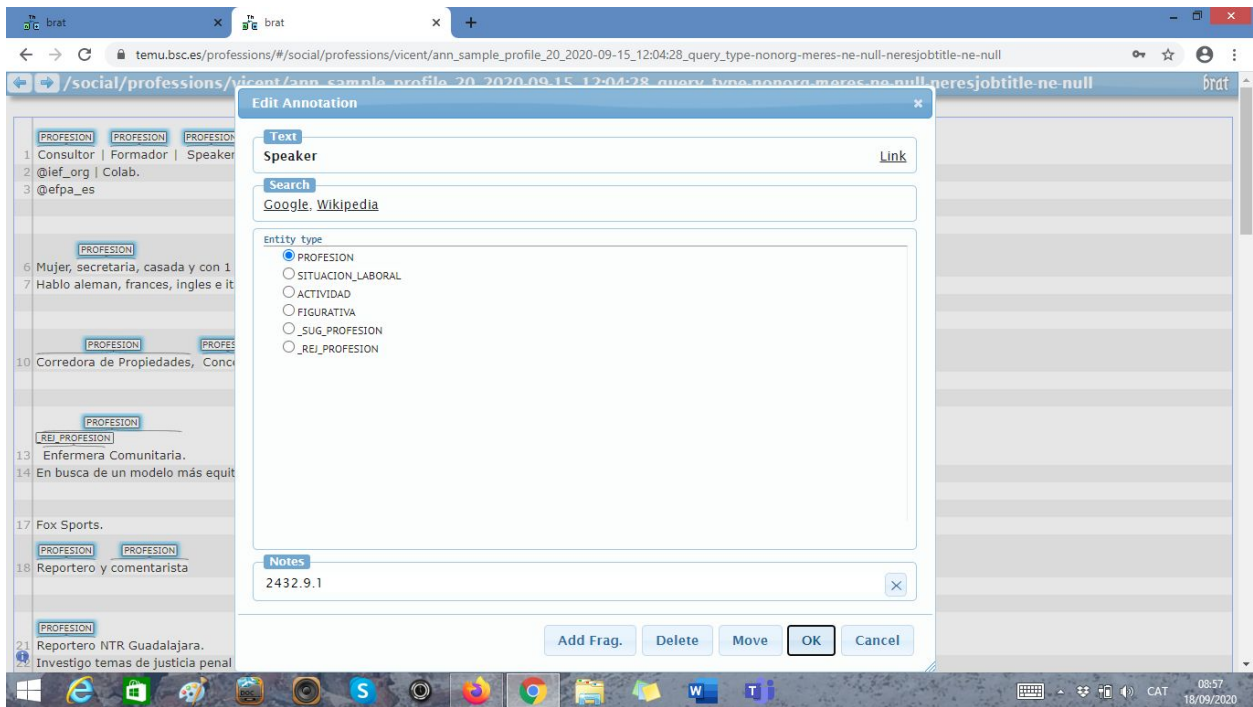


Figura 2: Interfaz de anotación de brat.

A continuación, se describen las reglas de anotación del corpus. La Tabla 3 reseña las reglas más generales, tales como el alcance de las anotaciones. La Tabla 4 describe las reglas positivas (qué casos deben anotarse), mientras que la Tabla 5 describe las negativas (qué no debe anotarse). Por último, la Tabla 6 señala algunos casos especiales que pueden encontrarse en el corpus.


A modo de aclaración, cuando se dude de si la mención corresponde a una profesión, deben consultarse principalmente los recursos mencionados (clasificación ESCO, etc.). Por ejemplo, el código 5161 de ESCO contempla *Astrólogos, adivinadores y afines*.

TABLA 3. REGLAS GENERALES, ORTOGRÁFICAS Y TIPOGRÁFICAS

Número e identificación	Regla	Ejemplo
G1 Recursos <i>resources</i>	Uso de recursos de conocimiento externos (limitar a profesiones excepto cuando sea imprescindible).	ESCO, SNOMED CT, DeCS, WordNet, diccionarios, www
G2 Claridad <i>clear</i>	No anotar menciones dudosas aún después de consultar las guías.	
G3 Actualizar <i>report</i>	Reportar casos especiales para mejorar y completar las guías de anotación.	<i>Code Monkey</i>
G4 Independencia <i>independent</i>	Las anotaciones realizadas sobre un texto (tuit o caso clínico) deben realizarse de forma independiente, sin tener en cuenta textos anotados previamente.	
G5 Tokens <i>5_tokens</i>	Límite de mención: una mención podrá contener un máximo de 5 tokens. Estos tokens tienen que estar en la misma frase.	<i>Asesor de Materiales de Aislamiento para Eficiencia Energética</i> (esta última parte no se puede incluir) <i>la infame "Fiscalía General del Gobierno"</i> <i>Profesor Asociado UAL-Dpto. Economía y Empresa</i> <i>periodista de El Correo, recientemente ascendido al</i>




		<i>departamento de deportes</i> (entre periodista y deportes hay 9 tokens)
G6 Relevancia <i>relevance</i>	Se anotará el mínimo texto significativo para denotar la ocupación, siempre que no pase del máximo de tokens.	<i>el alcalde de Madrid</i> <i>el alcalde de Madrid</i> <i>director de @CMNUCC</i> <i>el subsecretario de Prevención y Promoción de la salud, Hugo López</i> <i>Directora área psicología</i> <i>acredita a los trazadores de contagios</i> <i>abogada de oficio</i> <i>pediatra de urgencias</i>
G7 Solo completos <i>complete</i>	Las menciones solo pueden contener tokens completos.	<i>Especialista en Enfermería Familiar y Comunitaria</i> <i>Especialista en Enfermería Familiar y Comunitaria</i>
G8 Negaciones <i>negations</i>	Las ocupaciones negadas en el texto también se anotan.	<i>ni filósofo ni ministro</i> <i>Escribo, pero no soy escritor</i>
G9 Errores <i>mistakes</i>	Se anotarán también menciones con errores ortográficos y tipográficos.	<i>Ingenero de Caminos, Canales y Puertos</i>

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 9/27

G10 Sin vacíos <i>no_empty</i>	Las menciones NO pueden comenzar ni acabar con espacios en blanco.	<i>Padre y profesional de la Industria Farmacéutica</i> <i>Padre y profesional de la Industria Farmacéutica</i>
G11 Sin puntuación <i>no_puntuación</i>	NO incluir como parte de las menciones los símbolos de puntuación que no sean imprescindibles para la mención.	<i>técnico de proyectos,</i> <i>técnico de proyectos</i>
G12 Solapamiento <i>overlap</i>	Solapamiento de menciones y entidades anidadas.	Médico de Familia y Especialista en Medicina Interna, anotaremos primero <i>Médico de Familia</i> (médico de familia, primera profesión) Después, anotaremos <i>Médico de Familia y Especialista en Medicina Interna</i> como una segunda profesión (médico especialista)

TABLA 4. REGLAS POSITIVAS PARA LA ANOTACIÓN DE PROFESIONES/OCUPACIONES


Número e identificación	Explicación	Ejemplo
P1 Todas las menciones <i>all_mentions</i>	Hay que anotar todas las menciones de ocupaciones, aunque se repitan o no se refieran al autor del tuit ni al paciente de la historia clínica.	Autor tuit: <i>Maestro de infantil</i> Texto tuit: <i>en esta ciudad el alcalde está fallando</i>
P2 Siglas y abreviaturas <i>acr_abbreviatio</i>	Se anotarán menciones en forma abreviada y con siglas en los casos:	<i>Especialista en RRHH</i> <i>Relaciones Publi. y madre de dos niños</i>

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 10/27

<i>ns</i>	1. que sean universalmente reconocibles; y 2. que el Buscador de Google las identifique sin problema.	CRF
P4 Presencia de @, # <i>at_numeral</i>	También hay que anotar las menciones que indiquen claramente ocupaciones incluso cuando vayan precedidas de @, #. [Excepción: ver regla N3]	Especialista en gestión de #eventosdeportivos @bombers _man El verdadero virus
P5 Todos los idiomas <i>all_languages</i>	Se anotarán las menciones en todos los idiomas reconocibles (castellano, catalán e inglés). El idioma por defecto será el español internacional, añadiremos al código CAT para catalán y EN para inglés.	Advocat , abogado University Lecturer in Economics
P6 Slang <i>all_slangs</i>	Anotaremos (y recogeremos) ocupaciones expresadas en lenguaje informal/«slang». En una tabla y el diccionario añadiremos su equivalente formal. Ver más en https://www.oxford-royale.com/articles/professional-slang/#a1d=4ad63106-3c36-41fd-b27a-2e362f6d4ab0 .	Code monkey (programador básico, estudios equivalentes educación secundaria FP). Ver más en https://urbanthesaurus.org/synonyms/computer%20programmer .
P7 Menciones figurativas <i>fig_mentions</i>	Anotaremos menciones de profesiones aunque sean	Cowboy <i>crepuscular</i> Maestro <i>de nada</i>




	claramente metafóricas o figurativas.	
P8 Presidente de entidad gubernamental <i>gov_ent_president</i>	Al anotar el cargo político <i>presidente</i> , se anotará también la entidad que presiden. Asimismo, anotaremos las entidades de cargos públicos si se trata de organismos nacionales o supranacionales reconocidos.	<i>presidente supremacista de la Generalitat</i> <i>presidente de este Gobierno</i> <i>Presidenta de la Comunidad de Madrid</i> <i>directora general de ONU Mujeres</i> <i>presidenta de la Comisión Europea</i> <i>presidente del FMI</i> <i>presidente de la Universidad</i> <i>presidente de la Universidad</i>
P9 Políticos <i>prof_politics</i>	Todos los cargos dentro de partidos políticos, sindicatos y ONG se considerarán profesiones.	<i>presidente de cultura del PP de Cádiz</i> <i>vicepresidente de la ONG Porteños Solidarios del Sur</i>
P10 Grupos profesionales <i>occ_groups</i>	Las menciones que por contexto impliquen grupos <i>específicos</i> de profesionales también se anotan, así como equipos profesionales y las profesiones que sean homónimas a sus organismos.	<i>personal investigador</i> <i>policía local</i> <i>el equipo oncológico desaconsejó el medicamento</i> <i>en profesionales sanitarios tras una</i> <i>decidle al gremio que deje de presionar</i> <i>la Comisión de Sanidad para</i>

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 12/27

		<i>preguntar</i> Los Grupos Parlamentarios intervienen las autoridades británicas gestión del Gobierno de la crisis
P11 Médicos «residentes» <i>res_doctor</i>	Los médicos residentes se considerarán como médicos especialistas.	Médico residente de Pediatría
P12 Deportistas y artistas <i>ath_art_mentions</i>	Los deportistas y artistas (véase la siguiente regla) llevarán por defecto la etiqueta ACTIVIDAD. La etiqueta PROFESION se adjudicará cuando lleven el nombre propio detrás de su actividad.	baloncestista Alberto Díaz futbolista Luis Hernández Soy ciclista y padre de dos niños profesora de IES y actriz
P13 Autores/Escritores <i>auth_writers</i>	Los autores y escritores se considerarán profesionales cuando se mencionen junto a un nombre propio, el título de una de sus obras, o en relación con algún premio artístico o literario.	l' escritora Mercè Rodoreda el guionista de Casablanca soy poeta y soñador escritor finalista en el último Premio Planeta autora de Drames Rurals
P14 Ocupaciones generales <i>coach_director</i>	Monitor/a, entrenador/a, empresario/a, fundador/a, empresario/a, director/a, consultor/a, formador/a, representante, inversor/a,	monitor de tiempo libre Entrenador y jugador de fútbol , Real Oviedo hasta morir Formador homologado vinos de Jerez




	inventor/a se considerarán una profesión por defecto.	
P15 Miembros de ocupación activa <i>cur_occ_member</i>	La palabra miembro se anotará cuando indique una profesión activa, no la pertenencia a gimnasios, clubs, etc.	Un miembro del ejecutivo salía a miembro del comité olímpico
P16 Ocupaciones activas en duda <i>cur_occ_doubt</i>	Las ocupaciones precedidas o seguidas de expresiones como <i>sustituto, interim, intento de, ocasional, proyecto de</i> se anotarán y codificarán como tales.	<i>intento de</i> arquitecto Guionista <i>ocasional</i> <i>Proyecto de</i> jurista <i>Interim</i> manager
P17 Activismo <i>activism</i>	Se anotará la palabra <i>activista</i> con la etiqueta ACTIVIDAD, pero no así el tipo de activismo en concreto (<i>feminista</i>), que cuando se encuentren, se considerarán como ideología y se omitirán. Terrorismo se considerará tipo de activismo y no se anotará.	activista <i>feminista</i> activista <i>feminista</i> Periodista y feminista
P18 Actividad probable <i>not_lik_activity</i>	Si hay tuits con profesiones o actividades múltiples, utilizaremos la etiqueta ACTIVIDAD en todas las que se refieran a actividades artísticas, deportivas, de participación comunitaria, aficiones, etc.	Actriz , activista <i>feminista</i> , <i>anarkotriana</i> , bvloguera , artista <i>multi-INDisciplinADA</i> , música , ama de casa y <i>enferma de odio con aspiraciones místicas</i>
P19 Prefijos	Los prefijos <i>co-</i> y <i>ex-</i> siempre se anotarán junto con la	ex-presidente

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 14/27

<i>prefix</i>	mención, pero no nos modificarán la codificación de la ocupación.	ex-presidente co-director exjuez
P20 Trabajadores o empleados <i>workers</i>	Las oraciones donde aparezcan los términos <i>empleada/o en, trabajador/a en, persona del se</i> anotarán como profesión. Nota: <i>trabajador</i> y derivados se anotarán solo en su forma sustantiva.	<i>empleada en carpintería mecánica</i> <i>trabajador en canteras</i> <i>persona del ámbito sanitario</i>
P21 Sinónimos <i>synonyms</i>	Se anotarán sinónimos de profesiones expresados como: <i>profesión:cocinar; ocupación:hogar.</i>	<i>Paciente femenino de 68 años de edad, diestra, casada, ocupación: hogar</i> (sinónimo de ama de casa)
P22 Expertos o especialistas <i>expert_specialist</i>	Anotaremos menciones que incluyan la palabra <i>experta/o en o especialista en</i> si la especialidad se menciona claramente dentro del número máximo de tokens.	<i>profesora experta en pedagogía lúdica</i> <i>física experimental especialista en biomateriales</i> <i>especialista en Estados Unidos y Relaciones Multiculturales</i>

TABLA 5. REGLAS NEGATIVAS PARA LA ANOTACIÓN DE PROFESIONES/OCUPACIONES

Número e identificación	Explicación	Ejemplo
N1 Títulos académicos <i>no_aca_degrees</i>	No se incluirán calificaciones académicas, a no ser que impliquen una ocupación activa, a excepción del <i>Doctor/a en Medicina y Cirugía.</i>	<i>Doctor en Historia</i> <i>Máster en Bioquímica</i>


	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 15/27

		<i>Licenciada en Medicina y Cirugía</i> <i>doctor en Medicina y Cirugía</i> <i>doctorando en procesamiento de aguas (TAG Actividad)</i>
N2 Solo profesiones <i>no_orgs</i>	No se incluirán centros de trabajo ni organizaciones, a no ser que indiquen la profesión. En este caso limitaremos la anotación a lo requerido para la profesión.	<i>Ayuntamiento y Centro de Salud unidos contra el virus</i> <i>CEO de Laboratorios Almirall</i>
N3 Profesiones contenidas <i>no_occ_contained</i>	No se anotarán menciones de profesiones contenidas dentro de un mismo token (palabra).	<i>@elEconomistaes</i> <i>@Economista</i>
N4 Disciplinas <i>no_disciplines</i>	No se anotan menciones impersonales o expresiones verbales que denotan profesión.	<i>Carrera militar</i> <i>Psicología</i> <i>Investigando en salud</i>
N5 Profesiones ambiguas <i>no_ambiguous</i>	Sin contexto que denote profesión, las palabras ambiguas <i>modelo, defensor, segundo, padre, responsable, líder, jefe, organizador, agente</i> no se anotarán.	<i>organizador de Barbacoas</i> <i>organizadora de congresos</i> <i>responsable regional de sanidad</i>
N6 Connotaciones <i>no_connotations</i>	Las connotaciones de deporte no explícitas no se anotarán.	<i>campeona del mundo de pádel</i> <i>Jugador de bádminton</i>
N7 Adjetivos <i>no_adjectives</i>	No se anotarán adjetivos aunque sean igual que el sustantivo de una profesión.	<i>el estudio médico demostró</i> <i>era muy trabajadora</i>


N8 Solo humanos <i>only_humans</i>	Las posibles menciones de profesiones que no se refieran a seres humanos no se anotarán.	<i>drones socorristas</i> <i>perros guía</i>
N9 Direcciones <i>no_address</i>	Cuando el nombre de la profesión aparezca en una dirección (calle, avenida...) u otro sitio geográfico no se anotará.	Calle <i>Alcalde Tárrega</i> C/ <i>Almirante Churruca</i> <i>Hospital Obispo Polanco</i>
N10 Solo sinónimos <i>only_synonyms</i>	Los grupos de profesionales que no sean sinónimos a la profesión no se anotarán.	<i>decidle al gremio que deje de presionar</i> <i>la Comisión de Sanidad para preguntar</i> <i>Los Grupos Parlamentarios intervienen</i> <i>gestión del Gobierno de la crisis</i> <i>la policía se presentó</i> <i>la guardia civil acudió al lugar del accidente</i>
N11 Nobleza <i>no_nobility</i>	Los títulos nobiliarios no se anotarán, excepto cuando equivalgan a jefe de estado.	<i>Princesa de Asturias</i> <i>la reina de Inglaterra</i> <i>el copríncep d'Andorra</i>

TABLA 6. CASOS ESPECIALES DE ANOTACIÓN


Número e identificación	Explicación	Ejemplo
-------------------------	-------------	---------

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 17/27

E1 Doctores <i>occ_doctors</i>	Por convención, asumimos que las expresiones <i>doctora/doctor</i> (también <i>Dra., Dr.</i>) se refieren a médicos.	La <i>doctora</i> no me avisó de los temblores
E2 Profesional <i>pro_modifier</i>	El adjetivo <i>profesional</i> se añadirá cuando sirva para desambiguar profesión de actividad.	<i>Modelo profesional</i> <i>Profesional administrador financiero</i>
E3 Blogueros <i>blog_activity</i>	<i>Blogger, vlogger (youtuber, tiktoker, instagramer, streamer)</i> y sus variantes se considerarán profesiones. Se añadirá la N al código cuando se considere apropiado.	<i>Blogero de #gamificacion</i> <i>El instagramer artístico de Hospitalet</i>
E4 Conductores <i>occ_drivers</i>	La palabra <i>conductor</i> se interpretará por defecto como conductor de vehículo. En algunas ocasiones se tendrá que desambiguar.	<i>Conductor de autocar</i> <i>Conductor de programa radial</i> <i>Conductor eléctrico</i>
E5 Dueños <i>occ_owners</i>	La palabra <i>dueña/o, propietaria/o</i> en todas sus variantes de género y número se anotarán si se refiere a establecimientos de hospitalidad.	<i>dueño del bar</i>
E6 Realeza <i>even_emeritus</i>	Las menciones de la realeza se anotarán, junto con el adjetivo <i>emérito</i> cuando esté incluido.	la <i>reina Isabel II</i> el <i>rey emérito</i>
E7 Guías oficiales <i>occ_guides</i>	La palabra <i>guía</i> solo se anotará cuando el contexto denote profesión.	<i>Guía turístico</i> <i>Guía telefónica</i>

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 18/27

E8 Primera dama <i>first_lady</i>	<i>Primera dama</i> se considerará actividad.	<i>Saben el nombre de la primera dama</i>
E9 Profesor particular <i>prof_part</i>	<i>Profesor particular</i> se considerará una profesión al ser un término acuñado y ampliamente reconocido como una ocupación.	<i>ejerzo de profesor particular contacta</i>

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 19/27


3. CODIFICACIÓN

En un segundo paso, normalizaremos las profesiones o situaciones laborales que hayamos encontrado. Para ello, utilizaremos distintas ontologías y diccionarios (ESCO, SNOMED CT, otros), que habrá que conocer bien para adjudicar los códigos más apropiados. Estos códigos deben introducirse en la sección «Notes» de las anotaciones en la herramienta brat, junto a cualquier marca de idioma o semántica (estas marcas se explican más abajo en la Tabla 9). En el caso de que haya que introducir más de un código —o un código junto a una marca—, deben separarse usando una barra vertical (|) y sin espacios.

De manera general, la codificación se basará en la clasificación ESCO de ocupaciones⁴, que a la vez está basada en (y contiene «mapping» hacia) ISCO-08. Los grandes grupos de profesiones son los siguientes:

- 0 - Ocupaciones militares
- 1 - Directores y gerentes
- 2 - Profesionales científicos e intelectuales
- 3 - Técnicos y profesionales de nivel medio
- 4 - Personal de apoyo administrativo
- 5 - Trabajadores de los servicios y vendedores de comercios y mercados
- 6 - Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros
- 7 - Oficiales, operarios y artesanos de artes mecánicas y de otros oficios
- 8 - Operadores de instalaciones y máquinas y ensambladores
- 9 - Ocupaciones elementales

⁴ Disponible en: <https://ec.europa.eu/esco/portal/occupation> (última visita: 19/11/2020)

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 20/27

Crearemos una categoría de profesiones nuevas que no se han incorporado aún al ESCO, por ejemplo *youtuber* y *gamer*. Además, crearemos un diccionario inicialmente trilingüe (castellano, català, English) de profesiones.


A la hora de anotar el código vamos a utilizar el más granular posible. Por ejemplo, una *anestésista* se normalizará con el código de «médico especialista» (ESCO 2212.1), aunque el código correspondiente a «médico» también fuera correcto (ESCO 221). Otro ejemplo es el caso de *enfermera especializada en cuidados paliativos*. Asignaremos el código 2221.3 específico para «enfermera especialista», aunque el código 22 (Profesionales de la Salud) fuera también correcto. Sin embargo, si la especificidad de la anotación no se puede traducir en un código, anotaremos el término más general.

En algunos casos no podremos adjudicar una codificación sin tener en cuenta el contexto. En este caso, el contexto permitido será todo el contenido del tuit. En el ejemplo de la Figura 3, no nos sería posible adjudicar un código ESCO ni al *consultor especialista* ni al *manager* sin el contexto.



Figura 3: Ejemplo de codificación para la que es necesaria el contexto más amplio de la frase.

Es imprescindible que la persona anotadora conozca la clasificación ESCO, ya que muchas veces es difícil encontrar un término de búsqueda exacto. No obstante, pueden aparecer sinónimos que también se anotarán. Por ejemplo: el código 5412.1 de ESCO contempla *agente de policía*, pero también una sección llamada «alternative label» donde aparece *agente de la guardia civil* y *agente de policía local*, entre otros sinónimos (véase la Figura 4).

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 21/27

Spanish (es) 

agente de policía

[Discuss this topic in the Online Forum](#) 

Code

5412.1

Description

Los agentes de policía utilizan métodos de investigación para prevenir la delincuencia, perseguir y detener a los delincuentes y proteger al público de las actividades violentas y delictivas. Llevan a cabo actividades de vigilancia y patrulla para prevenir los delitos y detener a los delincuentes. Prestan apoyo a las víctimas y al público en general y cumplen funciones administrativas.

Scope notes

Includes state police officer.

Alternative label

agente de la guardia civil

agente de policía nacional

agente de policía local


policía

Figura 4: Ejemplo de código que incluye términos alternativos.

En algunos casos, la anotación será más granular que la codificación. Hay que tener en cuenta que son dos procesos *independientes*. Una vez realizada la anotación lo más granular posible (manteniendo la regla del máximo de tokens), se busca el código más adecuado. Por ejemplo:

54	REI PROFESION	Doctor en Derecho.
55	PROFESION	Profesor de Derecho Constitucional de la Universidad de Córdoba.
56		El Viso de los Pedroches (1969).

Figura 5: Ejemplo de anotación más granular que la codificación (I)

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 22/27

PROFESION	
45	Diputado de Área Servicios Sociales e Igualdad de Oportunidades @Dip_Va.
46	Máster en Economía de la Cultura.

Figura 6: Ejemplo de anotación más granular que la codificación (II)

Por un lado, en la Figura 5, aunque la anotación incluya *Constitucional*, el código nos indicará *profesor universitario de derecho*. Por otro lado, en la Figura 6 la anotación nos indica las responsabilidades del *diputado*, pero el código solo nos indicará que la persona mencionada es un diputado.


En ocasiones, en ESCO no existe un código para la categoría que buscamos. Esto ocurre, sobre todo, con las categorías que marcamos como SITUACIÓN LABORAL. En el caso de utilizar SNOMED CT⁵, la cifra del código irá precedida de SCTID. Por ejemplo, la palabra *estudiante* no aparece en ESCO y habría que utilizar el código de SNOMED, categorizándola como *SCTID: 65853000*. Puesto que lo normal es usar códigos ESCO, no debe aclararse el origen de estos códigos como con SNOMED. En el caso de que resulte imposible asignar un código, añadiremos XX a la mención, por ejemplo, *influencer*.

La Tabla 7 muestra algunos de los casos más comunes en los que necesitaremos utilizar códigos SNOMED:

TABLA 7: CÓDIGOS SNOMED SITUACIÓN LABORAL

Término	Sinónimos	Código Snomed CT
Self-employed/autònoma/autónoma	independiente; freelancer	SCTID: 160906004
retired/jubilada/emérito	pensionista	SCTID: 105493001

⁵ Disponible en <https://browser.ihtsdotools.org/?perspective=full&conceptId1=360863008&edition=MAIN/SNOMEDCT-ES/2020-04-30&release=&languages=es,en> (última visita: 19/11/2020)


	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 23/27

on maternity leave/permís maternitat/ permiso de maternidad		SCTID: 224457004
on paternity leave/permís paternitat/ permiso de paternidad		SCTID: 224458009
unemployed/a l'atur/en paro	búsqueda activa; en busca de trabajo; desempleada	SCTID: 73438004
student/estudiant/estudiante		SCTID: 65853000
worker/treballadora/obrero/trabajadora/obrero	empleada	SCTID: 106541005
housewife/mestressa de casa/ama de casa	homemaker	SCTID: 444168002
House husband/mestre de casa/amo de casa	stay-at-home dad	SCTID: 444169005

Hay algunas ocupaciones emergentes que aparecen en ESCO y que incluimos en la Tabla 8. Se debe recordar, como expresan las reglas P6 y E3, que estos ejemplos deben además marcarse añadiendo |N después del código si se tratan de neologismos.

TABLA 8: OCUPACIONES EMERGENTES QUE YA TIENEN CÓDIGO ESCO

Término	Sinónimos	Código ESCO
Vlogger	videobloguera/o vloguera/o streamer Youtuber	2642.1.17


	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 24/27

	TikToker	
Blogger	bloguera/o	2642.1.1

Por último, la Tabla 9 reseña algunos de los casos especiales de codificación y la Tabla 10 recopila algunas de las codificaciones más frecuentes a modo de ayuda para el corrector.

TABLA 9: CASOS ESPECIALES DE CODIFICACIÓN

Concepto	Ejemplo	Código
<i>reina/re</i> en activo se codificarán como profesión	<i>Reina</i> <i>Máxima de los Países Bajos</i>	
<i>reina/re emérito</i> se codificarán como jubilados	El <i>rey emérito</i> está en Dubái	SCTID: 105493001
cuando una mención justifique doble codificación (una de profesión y otra de supervisión o gestión), anotaremos las dos	<i>decano de jueces</i> <i>jefe de cocina</i> <i>jefe de enfermedades infecciosas</i>	1112 2612.1 (personal directivo de la administración pública y juez) 13 3434.1.1 (directores y gerentes de producción y operaciones y jefe de cocina) 1342 2212.1 (directores de servicios de salud y médico especialista)
Atletas y deportistas	soy <i>periodista</i> y <i>runner</i> (3421) el <i>futbolista</i> Suárez (3421.1)	Cuando no quede claro el status, los atletas y deportistas recibirán el código ESCO 3421; los atletas/deportistas profesionales, llevarán el código 3421.1
Idiomas, Neologismos, Slang y profesiones ilegales	La Comunidad Valenciana cuenta con 200 <i>rastreadores</i> (N)	Cuando el idioma de la anotación no sea en español, añadiremos CAT para

	ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias	01/12/2020	Página 25/27

		catalán, EN para inglés. Cuando se trate de un neologismo, añadiremos N, I para profesiones ilegales (ladrón, traficante, esbirro...) y S para slang.
--	--	--

TABLA 10. CODIFICACIONES FRECUENTES

Profesión	Código
abogada, jurista	2611.1
actriz/actor	2655.1
alcalde	1111.3
artista	265
cantante	2652.3.2
cocinero	5120.1
concejal	1111.1
diputada	1111.4
enfermera (diplomada en enfermería)	2221.3
escritora	2641.4
estudiante	SCTID: 65853000
fotógrafa	3431.1
juez	2612.1



ProfNER: Identificación de Profesiones y Ocupaciones en Redes Sociales Relacionadas con la Salud - Guías de anotación

M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima y V. Briva-Iglesias

01/12/2020

Página 26/27

ministra	1111.2
Médico con especialidad	2212.1
médico, Dr, doctor (sin especialidad)	221
periodista	2642.1
profesor	23
psicóloga	2634.2
sanitarios/profesionales de la salud	22
voluntaria	SCTID: 160772005