



# ProfNER: Identification of Health-Related Professions and Occupations in Social Networks

Annotation guides

*Plan for the Promotion of Language Technologies*

[M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima] <sup>1</sup>and [V. Briva-Iglesias] <sup>2</sup>

---

<sup>1</sup> Text Mining Unit, BSC

<sup>2</sup> D-REAL, Dublin City University

## TABLE OF CONTENTS

---

1	Introduction	2
2	Annotation	4
	Brat labels (Table 2)	4
	General rules (Table 3)	7
	Positive rules (Table 4)	9
	Negative rules (Table 5)	14
	Special cases (Table 6)	16
3	Coding	19

## 1. INTRODUCTION

---

The purpose of this guide is to facilitate the detection, codification and classification of professions/occupations and work situations expressed in natural language (Spanish, Catalan, and English) and contained mainly in a social network context ("Social Media"). The detection and codification can later be used to determine occupational groups at greater risk of: 1) infectious diseases; 2) other acute and chronic diseases; 3) mental health disorders; 4) risk behaviours, etc.

To obtain a "Gold Standard" we will use semi-supervised machine learning methods, in which professional annotators will correct and code the automatic pre-annotation of approximately 10,000 tweets under expert supervision.

Variants of the languages mentioned in the tweets will be considered. For example, we will take into account that the Argentinian expression *conductor radial* is equivalent to *radio speaker* in Spanish (e.g., *barrister* and *lawyer* in British and American English). Besides, we will create a trilingual dictionary of professions that covers the greatest number of occupations as they are expressed in the social networks, adding neologisms such as *gamer*, which are absent in the most known classifications.


For the classification, coding and "mapping" of results we will mainly rely on (in order of priority)

1	ESCO	<a href="https://ec.europa.eu/esco/portal/occupation">https://ec.europa.eu/esco/portal/occupation</a>
2	SNOMED CT (social context/occupations)	<a href="https://browser.ihtsdotools.org/?perspective=full&amp;conceptId1=14679004&amp;edition=MAIN/SNOMEDCT-ES/2020-04-30&amp;release=&amp;languages=es,en">https://browser.ihtsdotools.org/?perspective=full&amp;conceptId1=14679004&amp;edition=MAIN/SNOMEDCT-ES/2020-04-30&amp;release=&amp;languages=es,en</a>

All occupations in ESCO are mapped exactly to an ISCO-08 code. Organisations such as the National Institute of Statistics use NCO-11, which falls within the ISCO-08 conceptual framework. Idescat, with the advice of TERMCAT<sup>3</sup>, uses the CCO-2011, the translation into Catalan of the National Classification of Occupations 2011 (CNO-2011).

---

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

	<b>ProfNER: Identification of Health-Related Professions and Occupations in Social Networks</b>		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima and V. Briva-Iglesias	01/12/2020	Page 3/26

The annotator and coder will have experience in annotation and/or translation and word processing, with knowledge of at least 2 languages, one of which is English. In addition, before annotating and coding, they will need to be familiar with the ESCO classification of occupations. Finally, the colour code used to mark the different types of examples in these guides is shown in Table 1:

**TABLE 1: CONVENTIONS IN THE GUIDES**

1	Colour code: <b>incorrect entry</b>
2	Colour code by TAGs: <b>Profession</b> ; <b>Activity</b> ; <b>Working Status</b> ; <b>Figurative</b>

## 2. ANNOTATION

For the annotation we will use the brat tool, in which we will have the following categories when annotating an occupation. Please note that the examples have been gathered originally in Spanish, and here we show an English localization colour-marked (completely culturally adapted proposals, as they would appear in English-speaking social networks):

**TABLE 2. BRAT TOOL LABELS**

<b>TAG</b>	<b>Includes</b>	<b>Examples</b>
<b>PROFESSION</b> Occupations that provide the person with a wage or financial livelihood	Conventional professions; new Internet-related professions; sex workers; traffickers and other illegal occupations.  In the case of the prefixes <i>ex-</i> and <i>co-</i> , we will include them and consider the mention as the same profession.	<i>journalist from Radio Amanecer</i> <i>Future biomedical engineer</i> <i>co-executive director</i>
<b>WORKING STATUS</b>	Retired; unemployed; student/PhD candidate/apprentice/in training; temporally fired; self-employed; on parental leave; slave; homeless/beggar; worker; professional (no additional information).	<i>PR Student</i> 92,454 <i>temporally fired</i> and 678,700 <i>unemployed</i> <i>the worker, administrative assistant, aged 41</i>

<b>ACTIVITY</b> Unpaid occupations	Domestic workers; unpaid caregivers; non-professional sportspeople or artists; unpaid community workers; activists; volunteers; gurus; gamers.	<i>Venezuelan housewife</i> <i>Skier and photographer</i> <i>president of the association of parents</i> <i>impulsive gamer</i>
<b>FIGURATIVE</b> Metaphorical use of occupation names		<i>Painter of summer cinema ceilings</i> <i>Twilight Cowboy</i> <i>President of the republic of my house</i>
<b>SUG_PROFES SION</b>	The automatic suggestions on the PROFESSION label appear as _SUG_PROFESSION. The annotator should change this label to PROFESSION (or WORKING STATUS, ACTIVITY or FIGURATIVE) if it is correct.	
<b>REJ_PROFES SION</b>	If the automatic suggestion is incorrect, the annotator must change it to _REJ_PROFESSION.	

As an example, two figures are shown below: the first shows the final result of annotating a tweet and, the second, the brat interface.

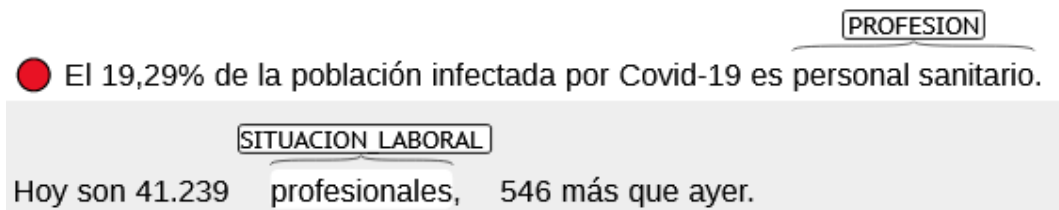


Figure 1: Final result of the annotation of a tweet.

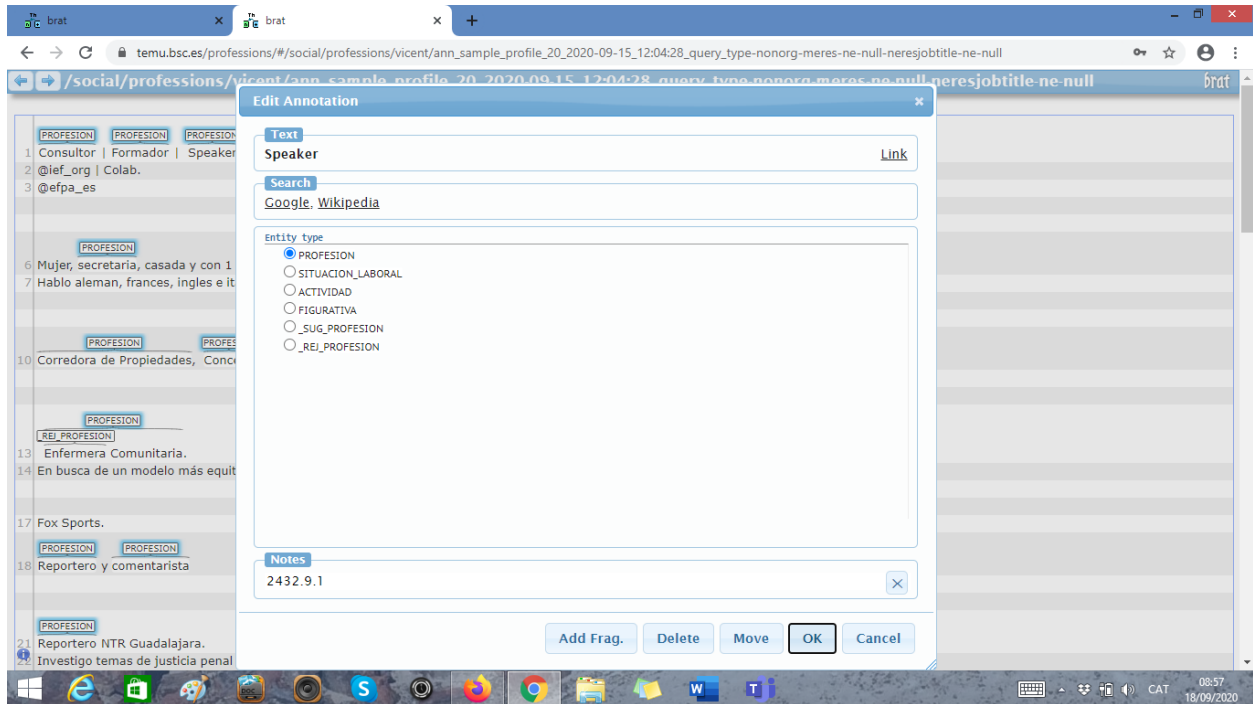


Figure 2: brat annotation interface.

The rules for corpora annotation are described below. Table 3 outlines the general rules, such as the scope of annotations. Table 4 describes the positive rules (what should be annotated), while Table 5 describes the negative rules (what should not be annotated). Finally, Table 6 points out some special cases that can be found in the corpus.

By way of clarification, when in doubt as to whether the mention corresponds to a profession, the resources mentioned should be consulted primarily (ESCO classification, etc.). For example, ESCO code 5161 covers *Astrologers, fortune-tellers and related workers*.

**TABLE 3. GENERAL, ORTHOGRAPHIC AND TYPOGRAPHICAL RULES**

<b>Number and identification</b>	<b>Rule</b>	<b>Example</b>
G1 Resources <i>resources</i>	Use of external knowledge resources (limited to occupations except where essential).	ESCO, SNOMED CT, DeCS, WordNet, dictionaries, www
G2 Clarity <i>clear</i>	Do not write down doubtful mentions even after consulting the guides.	
G3 Update <i>report</i>	Report special cases to improve and complete the annotation guidelines.	<i>Code Monkey</i>
G4 Independence <i>independent</i>	Annotations made on a text (tweet or health record) must be made independently, without taking into account previously annotated texts.	
G5 Tokens <i>5_tokens</i>	Mention limit: a mention may contain a maximum of 5 tokens. These tokens must be in the same sentence.	<i>Green Energy Efficiency</i> <i>Insulation</i> <i>Materials Advisor</i> (the previous part cannot be included, as there would be 6 tokens)  <i>the infamous "Government General</i> <i>Attorney</i>  <i>Associate Professor UAL Dep.</i> <i>Economics and Business</i>  <i>journalist for El Correo, recently</i> <i>promoted to the sports department</i>




		(between journalist and sports there are 9 tokens)
G6 Relevance <i>relevance</i>	The minimum significant text to denote occupation shall be annotated, provided it does not exceed the maximum number of tokens.	<i>the Mayor of Madrid</i> <i>the Mayor of Madrid</i> <i>director of @McDonald's</i> <i>director of @McDonald's</i> <i>Director of the Psychology Department</i> <i>public defender</i> <i>emergency pediatrician</i>
G7 Only complete <i>complete</i>	The mentions can only contain complete tokens.	<i>Family and Community Nursing Specialist</i> <i>Family and Community Nursing Specialist</i>
G8 Disclaimer <i>negations</i>	Occupations denied in the text are also annotated.	<i>neither philosopher nor minister</i> <i>I write, but I am not an author</i>
G9 Errors <i>mistakes</i>	Mentions with spelling and typographical errors will also be annotated.	<i>Engineer</i>
G10 No gaps <i>no_empty</i>	Mentions may NOT begin or end with blanks.	<i>Father and professional of the Pharmaceutical Industry</i> <i>Father and professional of the Pharmaceutical Industry</i>

G11 No score <i>no_punctuation</i>	DO NOT include as part of the mentions the punctuation symbols that are not essential for the mention.	<i>project technician,</i>  <i>project technician</i>
G12 Overlap <i>overlap</i>	Overlapping of mentions and nested entities.	In <i>Family Doctor and Specialist in Internal Medicine</i> , we will annotate Family <i>Doctor</i> first (family doctor, first profession) Afterwards, we will annotate <i>Family Doctor and Specialist in Internal Medicine</i> as a second profession (specialist doctor)

**TABLE 4. POSITIVE RULES FOR ANNOTATING PROFESSIONS/OCCUPATIONS**

<b>Number and identification</b>	<b>Explanation</b>	<b>Example</b>
P1 All the mentions <i>all_mentions</i>	All mentions of occupations must be annotated, even if they are repeated or do not refer to the author of the tweet or the patient in the medical record.	Tweet author: <i>Teacher</i>  Tweet text: <i>in this city the mayor is failing</i>
P2 Acronyms and abbreviations <i>acr_abbreviations</i>	Abbreviated mentions will also be annotated as acronyms in the following cases: <ol style="list-style-type: none"> <li>1. if they are universally recognisable; and</li> <li>2. if the Google Search engine identifies them without any problem.</li> </ol>	<i>HR Specialist</i>  <i>PR</i> and mother of two children  <i>CRF</i>
P4 Presence of @, # <i>at_numeral</i>	Annotators should also annotate every mention that clearly indicates an occupation,	<i>Specialist in management of #sportsevents</i>

	<b>ProfNER: Identification of Health-Related Professions and Occupations in Social Networks</b>		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima and V. Briva-Iglesias	01/12/2020	Page 10/26

	even if they are preceded by @, #. [Exception: see rule N3]	@firefighters_NY The real virus
P5 All languages <i>all_languages</i>	Mentions will be annotated in all recognisable languages (Spanish, Catalan and English). The default language will be international Spanish, we will add  CAT for Catalan and  EN for English to the code.	Advocat, lawyer University Lecturer in Economics
P6 Slang <i>all_slangs</i>	We will write down (and collect) occupations expressed in informal language/slang. In a table and the dictionary, we will add its formal equivalent. See more at <a href="https://www.oxford-royale.com/articles/professional-slang/#aId=4ad63106-3c36-41fd-b27a-2e362f6d4ab0">https://www.oxford-royale.com/articles/professional-slang/#aId=4ad63106-3c36-41fd-b27a-2e362f6d4ab0</a> .	Code monkey (basic programmer, equivalent studies to mandatory education). See more at <a href="https://urbanthesaurus.org/synonyms/computer%20programmer">https://urbanthesaurus.org/synonyms/computer%20programmer</a> .
P7 Figurative mentions <i>fig_mentions</i>	We will note mentions of professions even if they are clearly metaphorical or figurative.	Twilight Cowboy teacher of nothing
P8 President of a government entity <i>gov_ent_president</i>	When annotating the political occupation <i>president</i> , the entity they preside over shall also be annotated. Likewise, we will annotate the entities of public officers if they are working for recognized national or supranational bodies.	president of this government President of the Community of Madrid Director General of UN Women President of the IMF President of the University President of the University

<p>P9 Politicians <i>prof_politics</i></p>	<p>All positions within political parties, trade unions and NGOs will be considered professions.</p>	<p><i>President of Culture of the Democrats Party</i></p> <p><i>Vice-President of the NGO Porteños Solidarios del Sur</i></p>
<p>P10 Professional groups <i>occ_groups</i></p>	<p>Mentions that by context involve <i>specific</i> groups of professionals are also annotated, as well as professional teams and the professions that are homonymous with their bodies.</p>	<p><i>Research staff</i></p> <p><i>police</i></p> <p><i>the oncology team advised against the medicine</i></p> <p><i>in health professionals after a</i></p> <p><i>tell the union to stop lobbying</i></p> <p><i>the Health Commission ask</i></p> <p><i>The Parliamentary Groups intervene</i></p> <p><i>the UK authorities</i></p> <p><i>government management of the crisis</i></p>
<p>P11 Resident doctors <i>res_doctor</i></p>	<p>Resident doctors will be considered as specialist doctors.</p>	<p><i>Resident doctor in Paediatrics</i></p>
<p>P12 Sportsmen and artists <i>ath_art_mentions</i></p>	<p>Athletes and artists (see the following rule) will carry the ACTIVITY label by default. The PROFESSION label will be awarded when they bear their own name behind their activity, as this is an indicator of professionalism.</p>	<p><i>basketball player Alberto Díaz</i></p> <p><i>footballer Luis Hernández</i></p> <p><i>I am a cyclist and father of two children</i></p> <p><i>High-School teacher and actress</i></p>
<p>P13 Authors/Writers</p>	<p>Authors and writers are considered to be professionals</p>	<p><i>The writer Stephen King</i></p>

<i>auth_writers</i>	when the title of one of their works is mentioned alongside their own name or in connection with an artistic or literary prize.	<i>the <b>scriptwriter</b> of Casablanca</i> <i>I am a <b>poet</b> and a dreamer</i> <i><b>writer</b> finalist in the last Planeta Prize</i> <i><b>author</b> of The Lord of the Rings</i>
P14 General occupations <i>coach_director</i>	Monitor, trainer, entrepreneur, founder, businessperson, director, consultant, trainer, representative, investor, and inventor will be considered as profession by default.	<i><b>leisure monitor</b></i> <i><b>Coach</b> and <b>football player</b>. Barcelona until death</i> <i><b>Certified wine consumption trainer</b></i>
P15 Members of active occupation <i>cur_occ_member</i>	The word <i>member</i> will be annotated when it indicates an active profession, not membership of gyms, clubs, etc.	<i>A <b>member of the executive</b> left for</i> <i><b>member of the olympic committee</b></i>
P16 Active occupations in question <i>cur_occ_doubt</i>	Occupations preceded or followed by expressions such as <i>substitute, interim, attempted, occasional, draft shall be</i> annotated and coded as such.	<i>attempt to be an <b>architect</b></i> <i>Occasional <b>screenwriter</b></i> <i><b>Lawyer's</b> project</i> <i>Interim <b>manager</b></i>
P17 Activism <i>activism</i>	The word <i>activist</i> will be annotated with the label ACTIVITY, but not the specific type of activism ( <i>feminist</i> ), which will be considered as ideology and omitted. Terrorism will be considered a type of activism and won't be annotated.	<i>feminist <b>activist</b></i> <i><b>feminist activist</b></i> <i><b>Journalist</b> and <b>feminist</b></i>

<p>P18 <i>Probable activity not_lik_activity</i></p>	<p>If there are tweets with multiple professions or activities, we will use the ACTIVITY label in all those that refer to artistic, sports, community participation, hobbies, etc. This case is usual in social networks, and normally authors don't undertake all these occupations professionally.</p>	<p><i>Actress, feminist activist, anarchist, b/vlogist, multi-disciplinary artist, musician, housewife and hate sick with mystical aspirations</i></p>
<p>P19 <i>Prefixes prefix</i></p>	<p>The <i>co-</i> and <i>ex-</i> prefixes will always be annotated with the mention, but will not change the coding of the occupation.</p>	<p><i>co-director</i> <i>co-director</i></p>
<p>P20 <i>Workers or employees workers</i></p>	<p>Sentences where the terms <i>employee in, worker in, person in</i> be annotated as profession, when applicable. Note: <i>worker</i> and derivatives will be annotated only in their substantive form.</p>	<p><i>employee in mechanical carpentry</i> <i>worker in a quarry</i> <i>hard-worker</i> <i>person in the health field</i></p>
<p>P21 <i>Synonyms synonyms</i></p>	<p>Synonyms of professions expressed as: <i>profession:cooking;</i> <i>occupation:household.</i></p>	<p><i>68-year-old female patient, right-handed, married, occupation: household</i> (synonym for housewife)</p>
<p>P22 <i>Experts or specialists expert_specialist</i></p>	<p>We will annotate mentions that include the word <i>expert in</i> or <i>specialist</i> if the specialty is clearly mentioned within the maximum number of tokens.</p>	<p><i>expert in recreational pedagogy</i> <i>experimental physics specialist in biomaterials</i> <i>Specialist in United States and Multicultural Relations</i></p>

**TABLE 5. NEGATIVE RULES FOR RECORDING PROFESSIONS/OCCUPATIONS**


Number e identification	Explanation	Example
N1 Academic qualifications <i>no_aca_degrees</i>	Academic qualifications won't be included, unless they involve an active occupation, with the exception of the <i>Doctor of Medicine and Surgery</i> .	<i>Doctor of History</i> <i>Master in Biochemistry</i> <i>Degree in Medicine and Surgery</i> <i>Doctor of Medicine and Surgery</i> <i>PhD student</i> in water processing (Activity TAG)
N2 Professions only <i>no_orgs</i>	Workplaces and organisations won't be included, unless they indicate the profession. In this case we will limit the entry to what is required for the profession.	<i>Town Hall</i> and <i>Health Centre</i> united against the virus <i>CEO of Laboratories</i> Burton
N3 Professions contained <i>no_occ_contained</i>	No mention of professions contained within a token (word) will be annotated.	@the <i>Economist</i> @ <i>Economist</i>
N4 Disciplines <i>no_disciplines</i>	No impersonal mentions or verbal expressions denoting profession are annotated.	<i>Military career</i> <i>Psychology</i> <i>Health research</i>
N5 Ambiguous professions	Without context to denote profession, the ambiguous words <i>model, defender, second, parent,</i>	<i>Barbecue organizer</i> <i>congress organiser</i>

<i>no_ambiguous</i>	<i>responsible, leader, boss, organizer, agent</i> won't be annotated.	<i>regional health officer</i>
N6 Connotations <i>no_connotations</i>	Non-explicit sports connotations won't be annotated.	<i>paddle tennis world champion</i> <i>Badminton player</i>
N7 Adjectives <i>no_adjectives</i>	No adjectives will be annotated even if they are the same as the noun of a profession.	<i>the medical study showed</i> <i>she was a hard-worker</i>
N8 Human only <i>only_humans</i>	Possible mentions of professions that do not refer to human beings won't be annotated.	<i>guide dogs</i>
N9 Addresses <i>no_address</i>	When the name of the profession appears in an address (street, avenue...) or other geographical location it won't be annotated.	<i>Mayor Tárrega Street</i> <i>C/ Coronel Churruca</i> <i>Bishop Polanco Hospital</i>
N10 Only synonyms <i>only_synonyms</i>	Groups of professionals who are not synonymous with the profession won't be annotated.	<i>tell the union to stop lobbying</i> <i>the Health Commission to ask</i> <i>The Parliamentary Groups intervene</i> <i>government management of the crisis</i> <i>the police showed up</i>
N11 Nobility <i>no_nobility</i>	Noble titles won't be recorded, except when they are equivalent to head of state.	<i>Princess of Maryland</i> <i>the Queen of England</i> <i>the prince of Andorra</i>




**TABLE 6. SPECIAL CASES OF ANNOTATION**

<b>Number e identification</b>	<b>Explanation</b>	<b>Example</b>
E1 Doctors <i>occ_doctors</i>	By convention, we assume that the expressions doctor/doctor (also Dr.) refer to physicians.	<i>The <b>doctor</b> did not warn me about the tremors</i>
E2 Professional <i>pro_modifier</i>	The adjective <i>professional</i> will be added when it serves to disambiguate a profession from an activity.	<i><b>Professional model</b></i> <i><b>Professional financial administrator</b></i>
E3 Bloggers <i>blog_activity</i>	<i>Blogger, vlogger (youtuber, tiktokker, instagramer, streamer)</i> and their variants will be considered professions.  N will be added to the code when deemed appropriate.	<i><b>Blogger</b> of #gamification</i> <i>The artistic <b>instagramer</b></i>
E4 Drivers <i>occ_drivers</i>	The word <i>driver</i> will be interpreted by default as vehicle driver. In some cases, it will have to be disambiguated.	<i><b>Coach driver</b></i> <i><b>Electrical driver</b></i>
E5 Owners <i>occ_owners</i>	The word <i>owner, proprietor</i> in all its variants of gender and number shall be annotated if it refers to hospitality facilities.	<i><b>bar owner</b></i>
E6 Royalty <i>even_emeritus</i>	The mentions of royalty will be annotated, along with the adjective <i>emeritus</i> when included.	<i><b>Queen</b> Elizabeth II</i> <i>the <b>king emeritus</b></i>
E7 Official guides <i>occ_guides</i>	The word <i>guide</i> will only be annotated when the context denotes profession.	<i><b>Tourist guide</b></i> <i><b>Phone guide</b></i>

	<b>ProfNER: Identification of Health-Related Professions and Occupations in Social Networks</b>		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima and V. Briva-Iglesias	01/12/2020	Page 17/26

E8 First Lady <i>first_lady</i>	<i>First lady</i> will be considered as an activity.	<i>They know the name of the <b>first lady</b></i>
E9 Private teacher <i>prof_part</i>	<i>Private teacher</i> will be considered a profession as it is a term coined and widely recognized as an occupation.	<i>I work as a <b>private teacher</b></i>

	<b>ProfNER: Identification of Health-Related Professions and Occupations in Social Networks</b>		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima and V. Briva-Iglesias	01/12/2020	Page 18/26

### 3. CODING

---

In a second step, we will standardise the professions or working status we have encountered. To do this, we will use different ontologies and dictionaries (ESCO, SNOMED CT, others), which we will need to know well in order to assign the most appropriate codes. These codes should be entered in the "Notes" section of the annotations in the brat tool, together with any language or semantic markings (these markings are explained below in Table 9). If more than one code - or a code next to a mark - is to be entered, they should be separated using a vertical bar ( | ) and no spaces.

In general, the coding will be based on the ESCO classification of occupations<sup>4</sup>, which is based on (and contains "mapping" to) ISCO-08. The major groups of occupations are as follows:

- 0 - Armed forces occupations
- 1 - Managers
- 2 - Professionals
- 3 - Technicians and associate professionals
- 4 - Clerical support workers
- 5 - Service and sale workers
- 6 - Skilled agricultural, forestry and fishery workers
- 7 – Craft and related trades workers
- 8 - Plant and machine operators and assemblers
- 9 - Elementary occupations

---

<sup>4</sup> Available at: <https://ec.europa.eu/esco/portal/occupation> (last accessed 19/11/2020)

We will create a category of new professions that have not yet joined ESCO, for example *youtuber* and *gamer*. In addition, we will create an initially trilingual (Spanish, Catalan, English) dictionary of occupations.

When writing down the code we will use the most granular one that is possible. For example, an *anaesthetist* will be standardised with the code for “specialist physician” (ESCO 2212.1), even if the code for “physician” is also correct (ESCO 221). Another example is the case of a *nurse specialised in palliative care*. We will assign the specific code 2221.3 for “specialist nurse”, even if the code 22 (Health Professionals) is also correct. However, if the specificity of the notation cannot be translated into a code, we will note the more general term.

In some cases, we won’t be able to award a codification without taking into account the context. In this case, the permitted context will be the entire content of the tweet. In the example in Figure 3, it would not be possible for us to award an ESCO code to either the *specialist consultant* or the *manager* without the context.

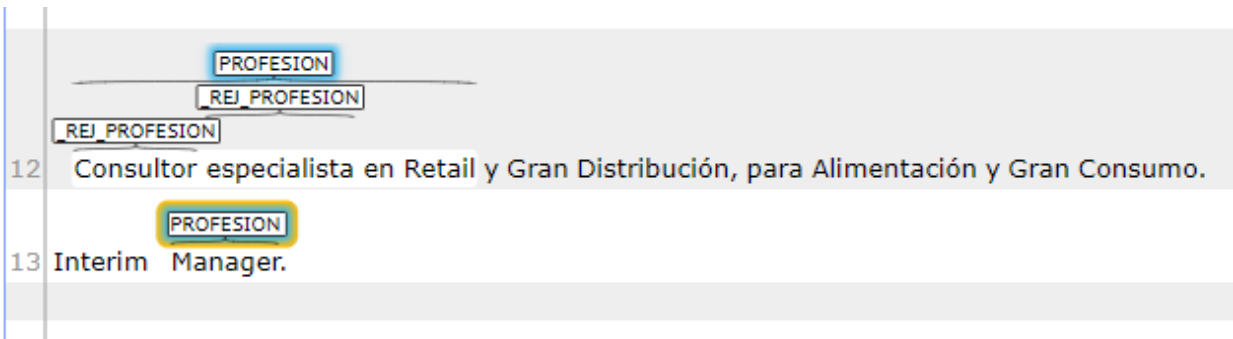





Figure 3: Example of coding for which the broader context of the sentence is necessary.

It is imperative for the annotator to know the ESCO classification, as it is often difficult to find an exact search term. However, synonyms may appear which will also be annotated and coded. For example: ESCO code 5412.1 includes a *police officer*, but also a section called “Alternative label”, where *local police officer* appears, among other synonyms (see Figure 4).

	<b>ProfNER: Identification of Health-Related Professions and Occupations in Social Networks</b>		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima and V. Briva-Iglesias	01/12/2020	Page 20/26

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

Figure 4: Example of a code that includes alternative terms

In some cases, the annotation will be more granular than the coding. It should be noted that these are two *independent* processes. Once the annotation is as granular as possible (maintaining the rule of maximum tokens), the most appropriate code is sought. For example:

54 REJ PROFESION  
Doctor en Derecho.

55 PROFESION  
Profesor de Derecho Constitucional de la Universidad de Córdoba.

56 El Viso de los Pedroches (1969).

Figure 5: Example of a more granular notation than coding (I)

45 PROFESION  
Diputado de Área Servicios Sociales e Igualdad de Oportunidades @Dip\_Va.

46 Máster en Economía de la Cultura.


	<b>ProfNER: Identification of Health-Related Professions and Occupations in Social Networks</b>		
	M. Agüero, E. Farré-Maduell, M. Krallinger, A. Miranda, S. Lima and V. Briva-Iglesias	01/12/2020	Page 21/26

Figure 6: Example of a more granular notation than coding (II)

On the one hand, in Figure 5, although the annotation includes *Constitucional*, the code will indicate a *university professor of law*. On the other hand, in Figure 6 the annotation indicates the responsibilities of the *Member of Parliament*, but the code will only indicate that the person mentioned is a Member of Parliament.

Sometimes, at ESCO, there is no code for the category we are looking for. This is especially the case with the categories we mark as WORKING STATUS. In the case of using SNOMED CT<sup>5</sup>, the code number will be preceded by SCTID. For example, the word *student* does not appear in ESCO and the SNOMED code should be used, categorising it as *SCTID: 65853000*. Since ESCO codes are normally used, the origin of these codes should not be clarified as with SNOMED. If it is impossible to assign a code, we will add XX to the mention, for example, *influencer*.

Table 7 shows some of the most common cases where we will need to use SNOMED codes:

**TABLE 7: SNOMED CODES FOR WORKING STATUS**

Term	Synonyms	Snomed CT code
Self-employed	independent; freelancer	SCTID: 160906004
retired/emeritus	pensioner	SCTID: 105493001
on maternity leave		SCTID: 224457004
on paternity leave		SCTID: 224458009

<sup>5</sup> Available at <https://browser.ihtsdotools.org/?perspective=full&conceptId1=360863008&edition=MAIN/SNO MEDCT-ES/2020-04-30&release=&languages=es,en> (last accessed 19/11/2020)

unemployed/a l'atur/en paro	active search; looking for work; unemployed	SCTID: 73438004
student/estudiant/estudiante		SCTID: 65853000
worker/treballador/trabajador	employee	SCTID: 106541005
housewife/housekeeper	homemaker	SCTID: 444168002
House husband	stay-at-home dad	SCTID: 444169005

It should be remembered, as expressed in the rules P6 and E3, that these examples should also be marked by adding 'N' after the code if they are neologisms.

**TABLE 8: POP-UP OCCUPATIONS THAT ALREADY HAVE ESCO CODE**

<b>Term</b>	<b>Synonyms</b>	<b>ESCO Code</b>
Vlogger	video-blogger vloguera/o streamer Youtuber TikToker	2642.1.17
Blogger	blogger	2642.1.1

Finally, Table 9 outlines some of the special cases of coding and Table 10 compiles some of the most frequent codings as an aid to the corrector.

**TABLE 9: SPECIAL CASES OF CODING**

Concept	Example	Code
active <i>queen/king</i> will be codified as a profession	<i>Queen</i> of the Netherlands	
<i>queen/king emeritus</i> to be codified as retired	The <b>King Emeritus</b> is in Dubai	SCTID: 105493001
where a mention justifies double coding (one of profession and another of supervision or management), we will use both	<i>dean of judges</i>  <i>head of kitchen</i>  <i>head of infectious diseases</i>	1112 2612.1 (senior public administration staff and judges)  13 3434.1.1 (production and operations directors and managers and head chef)  1342 2212.1 (health service managers and medical specialists)
Athletes and sportspeople	<i>i am a journalist and runner</i> (3421)  the <i>footballer</i> Suárez (3421.1)	Where the status is unclear, athletes and sportsmen will be given ESCO code 3421; professional athletes/sportsmen will be given code 3421.1
Languages, Neologisms, Slang and illegal professions	The Valencian Community has 200 <b>COVID-19 trackers</b> ( N)	When the language of the annotation is not in Spanish, we will add  CAT for Catalan,  EN for English. When it is a neologism, we will add  N,  I for illegal professions (thief, trafficker, henchman...) and  S for slang.



**TABLE 10. FREQUENT CODES**

<b>Profession</b>	<b>Code</b>
lawyer, jurist	2611.1
actress/actor	2655.1
Mayor	1111.3
artist	265
singer	2652.3.2
cook	5120.1
councillor	1111.1
Member of Parliament	1111.4
Nurse (graduate in nursing)	2221.3
writer	2641.4
student	SCTID: 65853000
photographer	3431.1
judge	2612.1
minister	1111.2
Specialist doctor	2212.1
doctor, Dr, doctor (no speciality)	221
journalist	2642.1

professor	23
psychologist	2634.2
health care/health professionals	22
volunteer	SCTID: 160772005