

The ProfNER shared task on automatic recognition of occupation mentions in social media: systems, evaluation, guidelines, embeddings and corpora

Antonio Miranda-Escalada

Barcelona Supercomputing
Center (BSC)
Barcelona, Spain

Eulàlia Farré-Maduell

Barcelona Supercomputing
Center (BSC)
Barcelona, Spain

Salvador Lima-López

Barcelona Supercomputing
Center (BSC)
Barcelona, Spain

Luis Gascó

Barcelona Supercomputing
Center (BSC)
Barcelona, Spain

Vicent Briva-Iglesias

SFI Centre for Research Training in Digitally-
Enhanced Reality, Dublin City University
Dublin, Ireland

Marvin M. Agüero-Torales

Barcelona Supercomputing
Center (BSC)
Barcelona, Spain

Martin Krallinger

Barcelona Supercomputing
Center (BSC)
Barcelona, Spain

{antonio.miranda, eulalia.farre, salvador.limalopez, lgasco, martin.krallinger}@bsc.es
vicent.brivaiglesias2@mail.dcu.ie, maguero@correo.ugr.es

Abstract

Detection of occupations in texts is relevant for a range of important application scenarios, like competitive intelligence, sociodemographic analysis, legal NLP or health-related occupational data mining. Despite the importance and heterogeneous data types that mention occupations, text mining efforts to recognize them have been limited. This is due to the lack of clear annotation guidelines and high-quality Gold Standard corpora. Social media data can be regarded as a relevant source of information for real-time monitoring of at-risk occupational groups in the context of pandemics like the COVID-19 one, facilitating intervention strategies for occupations in direct contact with infectious agents or affected by mental health issues. To evaluate current NLP methods and to generate resources, we have organized the ProfNER track at SMM4H 2021, providing ProfNER participants with a Gold Standard corpus of manually annotated tweets (human IAA of 0.919) following annotation guidelines available in Spanish and English, an occupation gazetteer, a machine-translated version of tweets, and FastText embeddings. Out of 35 registered teams, 11 submitted a total of 27 runs. Best-performing participants built systems based on recent NLP technologies (e.g. transformers) and achieved 0.93 F-score in Text Classification and 0.839 in Named Entity Recognition. Corpus: <https://doi.org/10.5281/zenodo.4309356>

1 Introduction

The number of social media users and content is rapidly growing, with over 700 million tweets posted daily (James, 2019). The Social Media Mining 4 Health (SMM4H) effort (Klein et al., 2020; Magge et al., 2021) attempts to promote the development and evaluation of NLP and text mining resources to extract automatically relevant health-related information from social media data. As social media content is produced directly by users at a global scale, a variety of application scenarios of medical importance have been explored so far, including the use of social media for pharmacovigilance (Nikfarjam et al., 2015), medication adherence, (Belz et al., 2019) or tracking the spread of infectious and viral diseases (Rocklöv et al., 2019; Zadeh et al., 2019).

The Spanish-speaking community on Twitter is large, exceeding 30 million (Tankovska, 2019), which motivated the implementation of text mining efforts for health-related applications, in particular on drug-related effects (Segura-Bedmar et al., 2015, 2014; Ramírez-Cifuentes et al., 2020).

One of the current challenges for health-related social media applications is to generate more actionable knowledge that can drive the design of intervention plans or policies to improve population health. This is particularly true for infectious disease outbreaks like the COVID-19 pandemic,

where certain occupational groups and subpopulations have been at higher risk, due to direct exposure to infected persons, a higher degree of social interaction, work-related travels to high-risk areas or mental-health problems associated to work-induced stress. Early detection and characterization of at-risk professions is critical to design and prioritize preventive and therapeutic measures or even vaccination plans.

To date, occupational text mining has been used in clinical narratives (Dehghan et al., 2016) and to explore injuries in the construction sector (Cheng et al., 2012), mainly in English. However, it has not yet been systematically used in social media and clinical content in Spanish.

To implement a central NLP component for occupational data mining, namely the automatic detection of occupation mentions in social media texts, we have organized the ProfNER (SMM4H 2021) shared task. In this article, occupation mentions are all those elements that indicate the employment information of a person. Within occupations, we have identified three main labels: (i) "profession", occupations that provide a person with a wage or livelihood (e.g., nurse); (ii) "activity", unpaid occupations (e.g., activist); and (iii) "working status" (e.g., retired).

Resources released for the ProfNER track included annotation guidelines in Spanish and English, a consistency analysis to characterize the quality and difficulty of the track, a large collection of manually annotated occupation mentions in tweets, as well as FastText word embeddings generated from a very large social media dataset.

We foresee that the occupation mention recognition systems resulting from this track could serve as a key component for more advanced text mining tools integrating technologies related to opinion mining, sentiment analysis, gender-inequality analysis, hate speech or fake news detection. Moreover, there is also a clear potential for exploring occupation recognition results for safety management, risk behavior detection and social services intervention strategies.

2 Task Description

2.1 Shared Task Goal

ProfNER focuses on the automatic recognition of professions and working status mentions on Twitter posts related to the COVID-19 pandemic in Spanish.

2.2 Subtracks

We have structured the ProfNER track into two independent subtracks, one related to the classification of whether a tweet actually does mention occupations or not, and another subtrack on finding the exact text spans referring to occupations.

This setting was decided due to the high class imbalance in social media data. Indeed, only 23.3% of the Gold Standard tweets contained mentions of occupations. Then, detecting relevant tweets (subtrack A) may help to detect the occupation mentions (subtrack B).

Subtrack A: Tweet binary classification. This subtrack required binary classification of tweets into those that had at least a single mention of occupations and those that did not mention any.

Subtrack B: NER offset detection and classification. Participants must find the beginning and the end of relevant mentions and classify them in the corresponding category.

2.3 Shared Task Setting

The ProfNER shared task was organized in three phases run on CodaLab¹:

Practice phase. The training and validation subsets of the Gold Standard were released. During this period, participants built their system and assessed their performance in the validation partition.

Evaluation phase. The test and background partitions were released without annotations. Participants had to generate predictions for the test and the background sets, but they were evaluated only on the test set predictions. This prevented manual annotations and assessed whether systems were able to scale to larger data collections. Each team was allowed to submit up to two runs.

Post-evaluation phase. The competition is kept alive on CodaLab. Interested stakeholders can still assess how their systems perform.

2.4 Evaluation: Metrics and Baseline

For Subtrack A, systems have been ranked based on F1-score for the positive class (tweets that contain a mention). For Subtrack B, the primary evaluation metric has been the micro-averaged F1-score. In this second track, a prediction was successful if its span matched completely the Gold Standard annotation and had the same category. Only PROFESSION (profession) and SITUACION LABORAL

¹<https://competitions.codalab.org/competitions/28766>

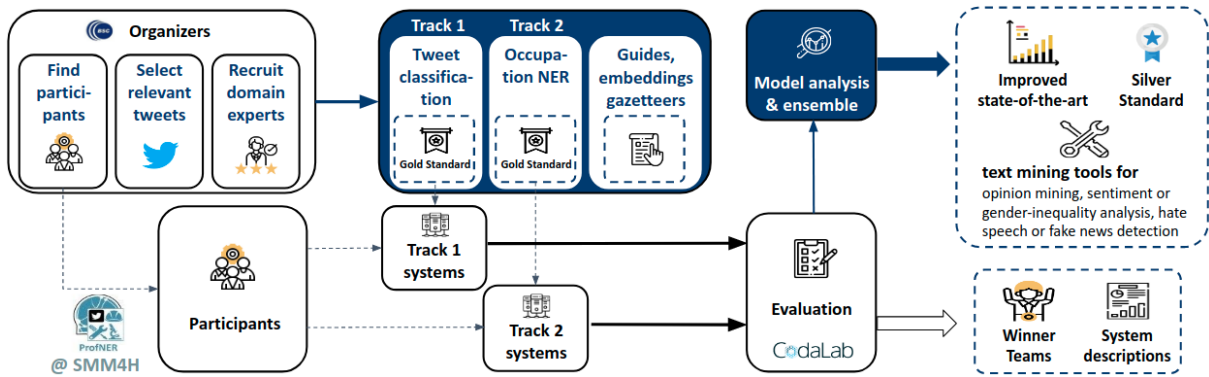


Figure 1: ProfNER Shared Task overview.

(working status) categories are considered in the evaluation.

Also, we have compared every system to a baseline prediction, a Levenshtein lexical lookup approach with a sliding window of varying length.

3 Corpus and Resources

3.1 ProfNER Gold Standard

The ProfNER Gold Standard is a collection of 10,000 COVID-related tweets in Spanish annotated with 4 types of occupation mentions: PROFESION (in English, "profession"), ACTIVIDAD ("activity"), SITUACION LABORAL ("working status") and FIGURATIVA (indicating that occupations are used in a figurative context).

The corpus has been split into training (60%), development (20%) and test (20%) sets. In addition, 25,000 extra tweets are provided without annotations as a background set. Table 1 contains an overview of the corpus statistics.

The corpus was carefully selected to include documents relevant to the challenges of the COVID-19 pandemic. It was obtained from a Twitter crawl that used keywords related to the pandemic (such as "Covid-19") and lockdown measures (like "confinamiento" or "distanciamiento", that are the Spanish translations of lockdown and social distancing), as well as hashtags such as "#yomequedoencasa" (#istayathome), to retrieve relevant tweets. Finally, we only kept the tweets that were written from Spain and in Spanish.

We filtered the tweets using the location attribute of the user profile and looked for the name of Spanish cities with more than 50K inhabitants, province names, autonomous region names, as well as any location specified simply as "Spain".²

²The list of place names was obtained from the Instituto

Gold Standard Quality. The corpus was manually annotated by expert linguists in an iterative process that included the creation of custom-made annotation guidelines, described in Section 3.3. Furthermore, we have performed a consistency analysis of the corpus: 10% of the documents have been annotated by an internal annotator as well as by the expert linguists. The Inter-Annotator Agreement (pairwise agreement) is 0.919.

	Documents	Annotations	Tweets with mentions	Tokens
Training	6,000	1,922	1,393	207,539
Validation	2,000	675	477	68,672
Test	2,000	636	463	68,713
Total	10,000	3,233	2,333	344,924
Background	25,000	81,922 (!)	7,394 (!)	805,961

Table 1: ProfNER corpus summary. (!) Background annotations are extracted from participants' predictions.

Gold Standard Format. Tweets were provided in plain UTF-8 text files, composed Unicode form. Every tweet is contained in a text file whose name is the tweet ID. The tweet classification annotations are distributed in a tab-separated file. The Named Entity Recognition (NER) annotations are distributed in Brat standoff format (Stenetorp et al., 2012) and in a tab-separated file (Fig. 2).

Translation to English. The ProfNER shared task attracted participants from many non-Spanish speaking countries. Besides, there exist more resources for social media text processing in English than in Spanish. For that reason, we have provided, as an additional resource, a machine-translated version of the ProfNER corpus. This will ease the

Nacional de Estadística (INE) <https://www.ine.es/dynt3/inebase/es/index.htm?padre=517&cap sel=525>

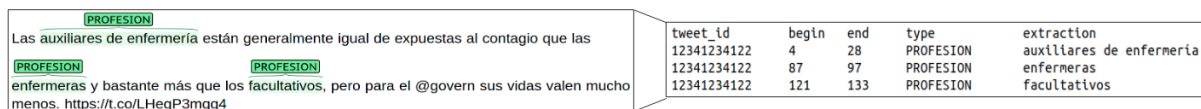


Figure 2: Example of an annotated tweet visualized with Brat annotation tool and its annotation file for track B.

comparison with systems working in English, provide support to participants working previously in English and explore the use of machine-translated corpora.

The complete ProfNER corpus —originally in Spanish— was translated into English by means of a state-of-the-art machine translation system based on recurrent neural networks.

The ProfNER Gold Standard and the English translation are distributed in Zenodo (Miranda-Escalada et al., 2020b).

3.2 ProfNER Silver Standard Corpus

The ProfNER test set was released together with an additional collection of 25,000 tweets as a background set. Participants have generated predictions for the test and background sets. The 81,922 predictions for this background set constitute the ProfNER Silver Standard corpus, similar to the ones of the CALBC (Rebholz-Schuhmann et al., 2010), the CodiEsp (Miranda-Escalada et al., 2020c) and the Cantemist (Miranda-Escalada et al., 2020a) initiatives. They will be released in the Zenodo Medical NLP community³.

3.3 ProfNER Guidelines

The creation of robust guidelines ensures dataset quality and replicability. Their main objectives are: (1) to capture all possible mentions of the entities of interest, especially occupations (*ex-trazadores de contagios*); and (2) to apply constraints to the mentions in order to obtain well-defined, replicable expressions (*ex-empleada en carpintería mecánica*).

ProfNER’s guidelines were created from scratch and iteratively refined to achieve maximum richness of mentions and until the Inter-Annotator Agreement was sufficiently high. All in all, six batches of annotations, corrections and reviews were required, reaching an agreement of 0.919. The final version includes 54 rules that describe the concepts to annotate and the associated constrictions. They are divided in four major groups: (i) 12 general rules, explaining the classification, orthographic and typographical aspects to be consid-

ered; (ii) 22 positive rules, explaining what should be deemed as an occupation; (iii) 11 negative rules, showing elements that should not be annotated; and (iv) 9 special cases of annotation. All rules are provided with illustrative corpus examples.

The guidelines were originally written in Spanish and later translated into English by a professional translator; both of them are freely available in Zenodo (Farré-Maduell et al., 2020).

3.4 ProfNER Embeddings

To our knowledge, we have released the first COVID-related embeddings in Spanish. They are especially suited for the ProfNER use case, since they are trained with 140 million COVID-related Spanish Twitter posts.

URL and mentions are substituted by the standard tokens URL and @MENTION, respectively. Embeddings were trained with FastText (Bojanowski et al., 2017) and the chosen embedding size was 300. CBOW and Skip-gram models in cased and uncased versions are available in Zenodo (Miranda-Escalada et al., 2021).

3.5 ProfNER occupations gazetteer

We have released the ProfNER gazetteer of occupations in Spanish, a resource that covers terminological resources from multiple terminologies (DeCS, ESCO, SnomedCT and WordNet) and occupations detected by Stanford CoreNLP in a large collection of social media Spanish profiles. The gazetteer can be found in Zenodo (Asensio et al., 2021).

4 Results

Participation Overview. Since this is the first task on the detection of occupational entities in social media, ProfNER has received considerable attention from the community. Indeed, 35 teams registered for the task (31 for the Subtrack A and 29 for the Subtrack B), and there were 27 submissions (15 in Subtrack A and 12 in Subtrack B) from 11 teams. Participant teams came both from the industry (3) and academia (8), and from different countries such as Spain, Portugal, Romania and China.

³<https://zenodo.org/communities/medicalnlp/>

Team	Country	A/I	Tool	Subtrack A			Subtrack B			Ref
				P	R	F1	P	R	F1	
Recognai	Spain	I	(2020)	0.93	0.93	0.93	0.84	0.84	0.84	(2021)
I2C	Spain	A	-	0.92	<u>0.92</u>	<u>0.92</u>	-	-	-	(2021)
Lasige-BioTM	Portugal	A	(2021)	0.95	<u>0.89</u>	<u>0.92</u>	0.81	0.66	0.73	(2021)
MIC-NLP	Germany	A&I	-	<u>0.95</u>	0.856	0.9	0.85	<u>0.80</u>	<u>0.82</u>	(2021)
RACAI	Romania	A	-	0.89	0.88	0.89	0.78	0.74	0.76	(2021)
GnaGna	Spain	I	-	0.93	0.78	0.85	-	-	-	-
UoB-NLP	UK	A	-	0.92	0.77	0.83	-	-	-	(2021)
SINAI	Spain	A	-	0.76	0.48	0.59	0.82	0.65	0.73	(2021)
Troy&AbedInTheMorning	Spain	A	(2021)	-	-	-	0.88	0.77	0.82	(2021)
CASIA_Unisound	China	A	-	-	-	-	<u>0.86</u>	0.64	0.73	(2021)
Baseline	Spain	A	(2021)	0.75	0.87	0.8	0.36	0.44	0.40	-

Table 2: Best result per team. Best result bolded, second best underlined. A/I stands for Academy/Industry.

System Results. Table 2 shows an overview of participants’ results. In the tweet classification subtrack, there are 3 systems with a similar performance (with F1-scores of 0.92 or 0.93), belonging to academia and industry. In the NER subtrack, the best performing system was developed by Recognai, an industry participant, with 0.839 F1-score. This system was based on a transformer architecture. The second best-performing system was from MIC-NLP, a partnership between Siemens AG and the Ludwig Maximilian University of Munich. They obtained a 0.824 F-score combining contextualized embeddings with BiLSTM-CRF. It is noteworthy that the best systems in terms of F1-score were also the systems with the highest recall, but not the highest precision.

Result analysis. Among the entities that ProfNER participants rarely detect, the proportion of SITUACION LABORAL is larger than in the entire corpus. Besides, some mentions with punctuation signs (particularly # or @) and capital letters are rarely predicted. For instance, "ministro del @interiorgob", "OFICIALES DE MESA" or "PENSIONISTAS" are never predicted. Notably, even though correct boundary detection remains a challenge for NER (Li et al., 2020), in our corpus entity length does not seem to influence predictability.

5 Discussion

To the best of our knowledge, ProfNER is the first occupational data mining effort in social media. It is also the first shared task on health applications of social media in Spanish. Specifically for the shared task, we have built a pioneer Gold Standard for Named Entity Recognition (NER) of occupations in social media in Spanish, the ProfNER corpus. It was generated following the ProfNER annotation guidelines that are shared with the community in

Spanish and in English. Finally, we have trained and released the first word embeddings for our use case (Twitter posts in Spanish related to COVID-19) and a gazetteer of relevant terms.

In addition, the ProfNER shared task can be used as template for future shared tasks on the recognition of occupations in social media. Indeed, the English translation of the annotation guidelines eases this research possibility.

ProfNER has aroused interest from both academia and industry. Interestingly, teams from non-Spanish speakers have participated in this task. Tweet classification systems reach high performances. However, the detection and classification of occupational data can still be improved.

We propose the use of the whole ecosystem of occupational text mining resources generated in this shared task (corpus, systems, guidelines, etc.) for building and evaluating novel systems that allow subpopulation characterization in social media in the current pandemic. This system has also the potential to assist public health policy makers in the prevention and management of current and future epidemics.

Moreover, beyond classical evaluation scenarios focusing on traditional quality evaluation using metrics like precision and recall, there is also a need to propose shared tasks and community benchmark efforts that do take into account involvement of end users, as was the case of the BioCreative interactive task (Arighi et al., 2011) or the technical evaluation on the robustness and implementation of named entity components (Leitner et al., 2008; Pérez-Pérez et al., 2016), especially when considering the volume and rapid change of social media data.

Acknowledgements

We acknowledge the Encargo of Plan TL (SEDIA) to BSC for funding of the ProfNER Shared Task. We would particularly like to thank the Social Analytics and the High Performance Artificial Intelligence groups of BSC who worked with us to obtain the Twitter corpus. Special mention to Carlos Rodriguez-Penagos, María Jose Rementeria and to Nataly Buslon Valdez. Finally, a big thank you to the rest of the SMM4H organizers for creating the perfect shared task ecosystem.

References

- Cecilia N Arighi, Phoebe M Roberts, Shashank Agarwal, Sanmitra Bhattacharya, Gianni Cesareni, Andrew Chatr-Aryamontri, Simon Clematide, Pascale Gaudet, Michelle Gwinn Giglio, Ian Harrow, et al. 2011. Biocreative iii interactive task: an overview. *BMC bioinformatics*, 12(8):1–21.
- Alejandro Asensio, Antonio Miranda-Escalada, Marvin M. Agüero-Torales, and Martin Krallinger. 2021. [Occupations gazetteer - ProfNER & MEDDOPROF - occupations, professions and working status terms with their associated codes](#). Funded by the Plan de Impulso de las Tecnologías del Lenguaje (Plan TL).
- Anja Belz, Richard Hoile, Elizabeth Ford, and Azam Mullick. 2019. Conceptualisation and annotation of drug nonadherence information for knowledge extraction from patient-generated texts. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 202–211.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.
- Ching-Wu Cheng, Sou-Sen Leu, Ying-Mei Cheng, Tsung-Chih Wu, and Chen-Chung Lin. 2012. Applying data mining techniques to explore factors contributing to occupational injuries in taiwan’s construction industry. *Accident Analysis & Prevention*, 48:214–222.
- Azad Dehghan, Tom Liptrot, Daniel Tibble, Matthew Barker-Hewitt, and Goran Nenadic. 2016. Identification of occupation mentions in clinical narratives. In *International Conference on Applications of Natural Language to Information Systems*, pages 359–365. Springer.
- Eulàlia Farré-Maduell, Salvador Lima López, Vicent Briva-Iglesias, Marvin M. Agüero-Torales, Antonio Miranda-Escalada, and Martin Krallinger. 2020. [Profner guidelines](#). Funded by the Plan de Impulso de las Tecnologías del Lenguaje (Plan TL).
- David Fidalgo, Daniel Vila-Suero, Francisco Aranda, and Ignacio Talavera. 2021. System description for profner - smmh: Optimized fine tuning of a pre-trained transformer and word vectors. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.
- Josh James. 2019. Data never sleeps 7.0. <https://www.domo.com/learn/data-never-sleeps-7>. Accessed: 2021-03-03.
- Ari Klein, Ilseyar Alimova, Ivan Flores, Arjun Magge, Zulfat Miftahutdinov, Anne-Lyse Minard, Karen O’connor, Abeed Sarker, Elena Tutubalina, Davy Weissenbacher, et al. 2020. Overview of the fifth social media mining for health applications (# smm4h) shared tasks at coling 2020. In *Proceedings of the Fifth Social Media Mining for Health Applications Workshop & Shared Task*, pages 27–36.
- Frances Laureano De Leon, Harish Tayyar Madabushi, and Mark Lee. 2021. UoB at ProfNER 2021: Data Augmentation for Classification Using-Machine Translation. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.
- Florian Leitner, Martin Krallinger, Carlos Rodriguez-Penagos, Jörg Hakenberg, Conrad Plake, Cheng-Ju Kuo, Chun-Nan Hsu, Richard Tzong-Han Tsai, Hsi-Chuan Hung, William W Lau, et al. 2008. Introducing meta-services for biomedical information extraction. *Genome biology*, 9(2):1–11.
- Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. 2020. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*.
- Arjun Magge, Ari Klein, Ivan Flores, Ilseyar Alimova, Mohammed Ali Al-garadi, Antonio Miranda-Escalada, Zulfat Miftahutdinov, Eulàlia Farré-Maduell, Salvador Lima López, Juan M Banda, Karen O’Connor, Abeed Sarker, Elena Tutubalina, Martin Krallinger, Davy Weissenbacher, and Graciela Gonzalez-Hernandez. 2021. Overview of the sixth social media mining for health applications (# smm4h) shared tasks at naacl 2021. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.
- J Alberto Mesa Murgado, Ana Belén Parras Portillo, Pilar López-Úbeda, M Teresa Martín-Valdivia, and L Alfonso Urena Lopez. 2021. Identifying professions & occupations in health-related social media using natural language processing. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.
- A Miranda-Escalada, E Farré, and M Krallinger. 2020a. Named entity recognition, concept normalization and clinical coding: Overview of the cantemist track for cancer text mining in spanish, corpus, guidelines, methods and results. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020)*, *CEUR Workshop Proceedings*.

- Antonio Miranda-Escalada. 2021. Bsc baseline. <https://github.com/tonifuc3m/profner-baseline>.
- Antonio Miranda-Escalada, Marvin M. Agüero-Torales, and Martin Krallinger. 2021. [Spanish covid-19 twitter embeddings in fasttext](#). Funded by the Plan de Impulso de las Tecnologías del Lenguaje (Plan TL).
- Antonio Miranda-Escalada, Vicent Briva-Iglesias, Eulàlia Farré, Salvador Lima López, Marvin M. Agüero-Torales, and Martin Krallinger. 2020b. [ProfNER corpus: gold standard annotations for profession detection in Spanish COVID-19 tweets](#). Funded by the Plan de Impulso de las Tecnologías del Lenguaje (Plan TL).
- Antonio Miranda-Escalada, Aitor Gonzalez-Agirre, Jordi Armengol-Estapé, and Martin Krallinger. 2020c. Overview of automatic clinical coding: annotations, guidelines, and solutions for non-english clinical cases at codiesp track of CLEF eHealth 2020. In *Working Notes of Conference and Labs of the Evaluation (CLEF) Forum*, CEUR Workshop Proceedings.
- Azadeh Nikfarjam, Abeed Sarker, Karen O’connor, Rachel Ginn, and Graciela Gonzalez. 2015. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association*, 22(3):671–681.
- Victoria Pachón Álvarez, Jacinto Mata Vázquez, and Juan Luís Domínguez Olmedo. 2021. Identification of profession & occupation in health-related social media using tweets in spanish. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.
- Martin Pérez-Pérez, Gael Pérez-Rodríguez, Obdulia Rabal, Miguel Vazquez, Julen Oyarzabal, Florentino Fdez-Riverola, Alfonso Valencia, Martin Krallinger, and Anália Lourenço. 2016. The markyt visualisation, prediction and benchmark platform for chemical and gene entity recognition at biocreative/chemdner challenge. *Database*, 2016.
- Vasile Pŕis and Maria Mitrofan. 2021. Assessing multiple word embeddings for named entity recognition of professions and occupations in health-related social media. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.
- Diana Ramírez-Cifuentes, Ana Freire, Ricardo Baeza-Yates, Joaquim Puntí, Pilar Medina-Bravo, Diego Alejandro Velazquez, Josep Maria Gonfaus, and Jordi González. 2020. Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. *Journal of medical internet research*, 22(7):e17758.
- Dietrich Rebbholz-Schuhmann, Antonio José Jimeno Yepes, Erik M Van Mulligen, Ning Kang, Jan Kors, David Milward, Peter Corbett, Ekaterina Buyko, Elena Beisswanger, and Udo Hahn. 2010. Calbc silver standard corpus. *Journal of bioinformatics and computational biology*, 8(01):163–179.
- Recognai. 2020. Recognai system code. <https://github.com/recognai/biome-text>.
- Joachim Rocklöv, Yesim Tozan, Aditya Ramadona, Maquines O Sewe, Bertrand Sudre, Jon Garrido, Chiara Bellegarde de Saint Lary, Wolfgang Lohr, and Jan C Semenza. 2019. Using big data to monitor the introduction and spread of chikungunya, europe, 2017. *Emerging infectious diseases*, 25(6):1041.
- Pedro Ruas. 2021. Lasige-biotm system code. <https://github.com/lasigeBioTM/LASIGE-participation-in-ProfNER>.
- Pedro Ruas, Vitor D. T. Andrade, and Francisco M. Couto. 2021. Lasige-biotm at profner: Bilstm-crf and contextual spanish embeddings for named entity recognition and tweet binary classification. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.
- Sergio Santamaría Carrasco and Roberto Cuervo Rosillo. 2021. Word embeddings, cosine similarity and deep learning for identification of professions & occupations in health-related social media. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.
- Sergio Santamaría. 2021. Troy&abendinthemorning system code. <https://github.com/ssantamaría94/ProfNER-SMM4H>.
- Isabel Segura-Bedmar, Santiago de la Peña González, and Paloma Martínez. 2014. Extracting drug indications and adverse drug reactions from spanish health social media. In *Proceedings of BioNLP 2014*, pages 98–106.
- Isabel Segura-Bedmar, Paloma Martínez, Ricardo Revert, and Julián Moreno-Schneider. 2015. Exploring spanish health social media for detecting drug effects. In *BMC medical informatics and decision making*, volume 15, pages 1–9. BioMed Central.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. 2012. Brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107.
- H. Tankovska. 2019. Countries with the most twitter users 2021. <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries>. Accessed: 2021-03-03.

- Usama Yaseen and Stefan Langer. 2021. Neural text classification and stacked heterogeneous embeddings for named entity recognition in smm4h 2021. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.
- Amir Hassan Zadeh, Hamed M Zolbanin, Ramesh Sharda, and Dursun Delen. 2019. Social media for nowcasting flu activity: Spatio-temporal big data analysis. *Information Systems Frontiers*, 21(4):743–760.
- Tong Zhou, Zhucong Li, Zhen Gan, Baoli Zhang, Yubo Chen, Kun Niu, Jing Wan, Kang Liu, Jun Zhao, Yafei Shi, Weifeng Chong, and Shengping Liu. 2021. Classification, extraction, and normalization : Casia_unisound team at the social media mining for health 2021 shared tasks. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.