

Received July 21, 2020, accepted August 6, 2020, date of publication August 24, 2020, date of current version September 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3019178

Understanding and Improving Disability Identification in Medical Documents

HERMENEGILDO FABREGAT¹, JUAN MARTÍNEZ-ROMO^{1,2},
AND LOURDES ARAUJO^{1,2}

¹Department of Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain

²Instituto Mixto de Investigación–Escuela Nacional de Sanidad (IMIENS), 28019 Madrid, Spain

Corresponding author: Hermenegildo Fabregat (gildo.fabregat@lsi.uned.es)

This work was supported by the Spanish Ministry of Science and Innovation through the DOTT-HEALTH Project under Grant PID2019-106942RB-C32, the PROSA-MED Project under Grant TIN2016-77820-C3-2-R, and the EXTRA-E-II Project under Grant IMIENS 2019.

ABSTRACT Disabilities are a problem that affects a large number of people in the world. Gathering information about them is crucial to improve the daily life of the people who suffer from them but, since disabilities are often strongly associated with different types of diseases, the available data are widely dispersed. In this work we review existing proposal for the problem, making an in-depth analysis, and from it we make a proposal that improves the results of previous systems. The analysis focuses on the results of the participants in DIANN shared task was proposed (IberEval 2018), devoted to the detection of named disabilities in electronic documents. In order to evaluate the proposed systems using a common evaluation framework, a corpus of documents, in both English and Spanish, was gathered and annotated. Several teams participated in the task, either using classic methods or proposing specific approaches to deal effectively with the complexities of the task. Our aim is to provide insight for future advances in the field by analyzing the participating systems and identifying the most effective approaches and elements to tackle the problem. We have validated the lessons learned from this analysis through a new proposal that includes the most promising elements used by the participating teams. The proposed system improves, for both languages, the results obtained during the task.

INDEX TERMS Disabilities, biomedical corpus, entity recognition, shared task analysis.

I. INTRODUCTION

The International Classification of Functioning, Disability and Health (ICF) defines a disability as any functional limitation that restricts, in any way, the capacity of a person to interact with the different environmental and personal factors that surround him or her. According to the World Health Organization (WHO), more than 15% of the world's population suffers from some form of disability. This organization also indicates that over 110 million adults have substantial difficulties in functioning. Disability rates are increasing due to an aging population, among other causes. Currently, due to the integration difficulties experienced by people with disabilities, different organizations and governments have specific integration plans on their agendas.¹

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato.

¹<https://www.un.org/development/desa/disabilities/news/news/sg-cosp.html> (Last visited: May 2020)

The application of Natural Language Processing (NLP) techniques in biomedical documents has facilitated significant advances in this area, allowing to improve information retrieval and knowledge inference processes. However, the automatic identification of mentions of disabilities has rarely been addressed in scientific works. Given its social impact and the little attention it has received, we decided to organize a shared task focused on identifying named disabilities in medical texts. The techniques developed for the task are useful to gather information on disabilities and to advance on their knowledge and prevention, as well as to help health institutions and governments to improve the social integration of the affected people.

Many evaluation campaigns or shared tasks are being organized in the field of natural language processing related to the biomedical domain [1]. They are extremely important to progress in the field, since the participating systems can compare their approaches in a fair way, using the same data and evaluation framework. In particular, some shared tasks

have focused on the detection of specific kinds of entities. The recognition of drugs in Spanish clinical documents has been addressed by tasks such as Pharmaconer [2], the extraction of drugs-drugs interactions has been worked in the DDI-Extraction 2013 task [3] and the recognition of named entities in clinical documents in French has been covered by labs such as CLEF eHealth 2015 [4].

As part of the 2018 IberEval workshop [5], we proposed the DIANN shared task focusing on the detection of disabilities in a manually annotated corpus provided to the participant teams. The corpus contains documents in both Spanish and English, and according to our knowledge, this is the first task organized to address this problem. Detecting disabilities involves specific problems, for several reasons. On the one hand, disability is a broad concept that can be interpreted in several ways. On the other hand, disabilities can be mentioned in free language and they are not constrained to a specific set of words. We took these particularities into account to define the task, as well as the annotation criteria adopted to prepare the corpus. The corpus is available at: <https://github.com/gildofabregat/DIANN-IBEREVAL-2018>.

This article aims to make a new proposal based on the analysis of the common and distinctive features used by the participating systems. From this analysis, along with the results obtained by the systems, we have drawn conclusions about the most influential features and how to combine them in an effective way. The lessons learned have allowed us to design a new proposal improving the results achieved by the participating systems.

This article is organized as follows: Firstly, we describe the corpus provided for the task and used as a common evaluation framework, including the followed methodology to collect the documents and the annotation criteria (Section III-A). Afterwards, we analyze and compare the different systems that were presented by each team (Section III-B). Extending this analysis, we highlight the most interesting elements and findings of each system (Section III-C). Finally, we present and evaluate a new approach that tries to capture the lessons learned from the shared task (Sections III-D and IV). We detail and discuss the main outcomes in the Sections V and VI.

II. RELATED WORK

Named Entity Recognition (NER) has been frequently considered in the biomedical domain as a critical task in extracting information from medical documents. It consists in identifying certain expressions of interest in documents and mapping them to the corresponding semantic categories (diseases, drugs, genes, etc.). Although some proposals for NER are based on natural language processing [6], most of them rely on machine learning techniques [7]–[9]. The techniques, which have been used in recent years, have particularly focused on deep learning [10]–[12]. Although there are systems that deal with several types of entities [6], [13], other works focus on specific types. In 2013, Segura-Bedmar *et al.* [3] proposed an evaluation task focused

on drug recognition and identification of adverse effects in scientific documents written in English. Works such as Gonzalez-Agirre *et al.* [2] have explored the detection of named drugs in Spanish documents. Gene recognition in specialized literature has been addressed by Yeh *et al.* [14], among others. Nevertheless, disabilities are not included as a specific semantic class to be identified, and these proposals do not distinguish a disability, usually a permanent condition, from other signs associated to diseases.

Orphanet,² the international database and portal on Rare Diseases (RD) and orphan drugs, is collecting information about the functional consequences associated to rare diseases [15]. It uses the Orphanet Functioning Thesaurus, derived and adapted from the International Classification of Functioning, Disability and Health — Children and Youth (ICF-CY [16]). Using the Orphanet database, in a previous work [17], we have collected the RDD corpus related to rare diseases and composed of scientific documents in English annotated with disabilities, as well as relationships between rare diseases and disabilities. We also developed a deep learning system for the detection of named disabilities and for the extraction of relationships between disabilities and rare diseases, obtaining remarkable results using Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN). Since detecting negation is paramount to understanding the meaning of texts correctly, the RDD corpus also includes negation and speculation (uncertainty about the mentioned facts) annotations.

Negation processing has been considered in highly relevant NLP tasks, such as sentiment analysis and relationship extraction [18], [19]. Many works about negation processing have been published for the English language [20], [21] although there are also other works that extend this study to other languages, such as Chinese [22], [23] or Spanish [24]. The most common approaches are rule-based [25]. However, there has also been proposals based on machine learning [26], and recently, specifically on deep learning [27].

III. MATERIALS AND METHODS

A. DIANN CORPUS

As far as we know, automatic disability annotation is a topic that has not been addressed for the Spanish language and barely for English [17]. An important outcome of the DIANN task has been the generation and release of a benchmark corpus used to evaluate all suggested proposals under the same criteria. The DIANN corpus consists of a collection of 1000 abstracts of scientific articles (500 published in English and 500 in Spanish). This collection was collected during 2017 and each abstract contains at least one mention of a disability. This corpus is very valuable since it includes abstracts of scientific papers provided in both languages (Spanish and English) and annotated under the same annotation criteria, allowing the analysis of systems focused

²<https://www.orpha.net/> (Last visited: May 2020)

on Spanish or English, and also systems focused on both languages.

Taking into account the definition of “disability” proposed by the ICF classification, and with the support of two medical doctors, the annotation process was carried out by three non-expert annotators. While in Spanish, this corpus gathers a set of 1555 mentions of disabilities (564 unique mentions), in English, 1656 disabilities have been annotated (583 unique mentions). Concerning the annotation process, the documents were gathered using queries related to a list of human functions and disabilities extracted from [17]. To summarize, a total of 45 different expressions were used to make these queries. Some of them are the following (English / Spanish):

- Blindness/Ceguera
- Paralysis/Parálisis
- Sight/Vista
- Cognitive/Cognitivo
- Intellectual/Intelectual
- Motor/Motriz

On the process of gathering and annotating, one document was considered as candidate for the corpus if it contained one of the searched expressions and if it was available in both languages (scientific papers with an abstract in both languages). In order to monitor and limit the number of retrieved documents for each searched expression, an iterative approach was applied where each iteration consisted in the retrieval and annotation of a maximum of 50 documents.

A definition of “disability” by the ICF classification was used as starting point for the annotation criteria. However, possible ambiguities had to be taken into account when considering a physical/mental condition or illness (due to its temporality or severity, for example) as a disability or not. To reduce the ambiguity effect, an additional premise was included in the annotation guidelines: “A disability is annotated, when it is assumed from the context that this condition is not of short duration or low severity.” Following guidelines from previous work [17], any possible modifiers around an annotation were considered part of it, e.g.:

The patient suffers from <dis> severe intellectual disability </dis> ... / El paciente sufre de <dis> discapacidad intelectual grave </dis>...

where the tags <dis> and </dis> refer to the beginning and end of a disability, respectively. In addition to annotations of disabilities, each document contains annotations of negations when they affect at least to one disability. For each annotated negation both, the scope and the associated triggers, are provided. The annotation criteria used to define the scope is inspired by [20]. The corpus contains a total of 62 negations in Spanish, and 63 negations in English.

Table 1 shows the inter-annotator agreement reached by type of annotation (negations or disabilities) and language (Spanish or English), and the number of annotations contained in the corpus. The inter-agreement indicates the percentage of agreement reached among the three annotators. Some documents do not contain the same amount of information for both languages, which explains the difference in the number of annotations per language. Even though there are differences in the agreement reached by the annotators for

TABLE 1. Corpus statistics for both languages. This table shows the total number of annotations and the agreement reached between the annotators.

Annotation	Language	Total	Inter-Agreement
Disability	English	1656	0.89
Disability	Spanish	1555	0.78
Negation	English	63	0.90
Negation	Spanish	62	0.87

each language, the achieved results indicate that the corpus is robust enough to be released and proposed as a benchmark. More details about the annotation format can be found in [28].

B. DIANN SHARED TASK

Using the DIANN corpus as benchmark, in the context of SEPLN³ 2018 conference and as part of the 2018 IberEval workshop, an evaluation task was organized to compile, discuss and share the knowledge of the participating teams regarding the automatic detection of named disabilities in scientific documents.

1) PARTICIPATING TEAMS

The eight participating teams presented quite promising approaches (18 runs for English and 19 for Spanish). This section analyses the proposals of all participating teams.

- **SINAI [29] - Group of Intelligent Systems of Information Access, University of Jaén.** SINAI group proposed an unsupervised system based on the generation of variants using UMLS (Unified Medical Language System) terminology and word embeddings. The system used two different biomedical entity extractors, MetaMap [30] for English and a similar tool for Spanish. After identifying potential expressions, the group proposed a filtering process based on two aspects: semantic categories manually identified as relevant and the analysis of the similarity between the candidate expressions and the expression “disability” using word embeddings. To address the detection of negated disabilities, SINAI used a bag-of-words based system to detect negation triggers and specific rules to determine the scope.
- **IxaMed [31] - Ixa Group, University of the Basque Country.** For the detection of named disabilities, this group used a deep learning system consisting of a Bidirectional Long Short Term Memory (Bi-LSTM) network and a Conditional Random Field (CRF) at the top. While to process the English documents, IxaMed used word embeddings and Brown clusters [32] as inputs, for the Spanish documents they only used word embeddings. Although IxaMed used different word embeddings depending on the language, both embeddings were generated using documents from the biomedical domain (English: MIMIC-III corpus [33] - Spanish: Electronic health records in Spanish). On the other hand, they used

³Sociedad Española de Procesamiento del Lenguaje Natural: <http://www.sepln.org>

a rule-based system, for both the detection of abbreviations and the detection of negation triggers. Finally, in order to detect the scope of the negation IxaMed group used a model based on neural networks.

- **IXA [34] - Ixa Group, University of the Basque Country.** This model deals with the detection of both disabilities and negation triggers using the *ixa-pipeline* [35] tool and the Perceptron implementation contained in the Apache OpenNLP project. Among the different types of inputs used by IXA were the public gazetteers and clusters, such as Brown and Clark [36]. In order to identify the scope of the negation, they assumed that each negation trigger identified in a sentence affects all disabilities found in that sentence. Given this assumption they set the scope from the first identified term in that sentence to the last.
- **GPLSIUA [37] - Natural Language Processing and Information Systems Group, University of Alicante.** Analogous to the proposal submitted by SINAI group, this approach consists of two parts: an expression extractor and a candidate expression filtering process. While the extraction method proposed by SINAI group is based on external biomedical resources, the method proposed by this team is much more generic, dealing only with the extraction of all noun phrases in each sentence. This proposal transfers the responsibility of filtering candidates to a machine learning system known as CARMEN [38], which uses Random Forest and is trained with several features of the text, e.g. suffixes, affixes, etc. This system also uses some contextual information on the “relevance” of lemmas of certain terms that appear in a fixed-size contextual window. Regarding the recognition of negated disabilities, this team proposes a dictionary-based system for the identification of negation triggers and the application of rules, such as those employed by IXA for the identification of the scope.
- **UPC_3 [39] - TALP research group, Polytechnic University of Catalonia.** Two different proposals were presented by this team: one was based only on CRF and other one was based on the tuple Bi-LSTM and CRF. While the CRF model was trained using several features, including lemmas, suffixes, part-of-speech and word embeddings, the Bi-LSTM+CRF model was trained only using word embeddings. UPC_3 carried out several experiments with different kinds of embeddings, one was trained with generic documents (sources such as Wikipedia among others) and another one trained only with documents from the biomedical domain (research articles, electronic health records, etc.). In order to avoid over-fitting during the training, this team proposed a semi-supervised methodology using unlabeled documents during the training. Finally, to process the negation and its scope they used a system known as ABNER [40].
- **UPC_2 [41] - TALP research group, Polytechnic University of Catalonia.** Based only on the use of a CRF for the identification of named disabilities, the system that proposed this team was trained with both syntactic and semantic features. Thus, some studied features include casing information such as capitalization and the use of non-alphanumeric elements, which are considered useful by different teams for the detection of abbreviations. This team also used a list of terms that were extracted from the training set in order to represent if a term found in the test set is of interest or not. Finally, UPC_2 used a NegEx-based [25] method to detect negated entities. After applying this method, this team used the distance between identified negation triggers and named disabilities to filter out possible false positives.
- **UC3M [42] - Human Language and Accessibility Technologies Group, Carlos III University.** This team used a Bi-LSTM+CRF based architecture to deal at the same time with the recognition of named disabilities and the identification of negation. Unlike other teams that used a similar architecture, UC3M trained the system using exclusively the following distributed representations: word embeddings for terms representation, character embeddings to represent n-grams of characters and sense embeddings for disambiguation. This team used a LSTM to process the sequence of character embeddings of each word. The output of this LSTM, along with word embeddings and sense embeddings, are the input of the Bi-LSTM+CRF model.
- **LSI_UNED [9] - NLP & IR group, National Distance Education University.** This team proposed an automatic annotation tool similar to UMLS MetaMap Transfer (MMTx) to extract biomedical concepts. The system generates different variants of the same disability aiming to improve coverage. This proposal uses external resources to perform some language processing tasks, and it begins with the thesaurus (lists of disabilities and body functions) processing in which the variants of the disabilities it contains are generated. Then, given a document, this system identifies the noun phrases and generates their variants. Variants of the disabilities and the body functions are generated in the document using Wordnet [43]. In addition, it is possible to configure the variant generation levels in the document and the thesaurus. In the last phase, the system employs a ranking based selection function which takes into account four measures: centrality, variation, coverage and cohesiveness.

2) SUMMARY OF EXPLORED FEATURES AND ALGORITHMS

Table 2 shows a comparison of the different features and resources used by each team to deal with the recognition of named disabilities. Both supervised and unsupervised systems are included in the comparison. Although the use of

TABLE 2. Comparison of features and resources used by participating teams. Each column represents one of the participants and each row represents if a feature or resource has been used or not. (G) Generic domain sources - (E) Specific domain sources - (C) Characters - (S) Sense.

	IXA	IxaMed	UPC_3	LSI_UNED	GPLSIUA	SINAI	UPC_2	UC3M
Embeddings		E	G-E			G		G-C-S
Clustering	X	X	X					
Dictionaries	X			X		X		
Part-of-Speech			X		X		X	
Wordnet				X				
Char n-grams	X				X			
Lemmas			X		X		X	
Casing		X	X	X			X	

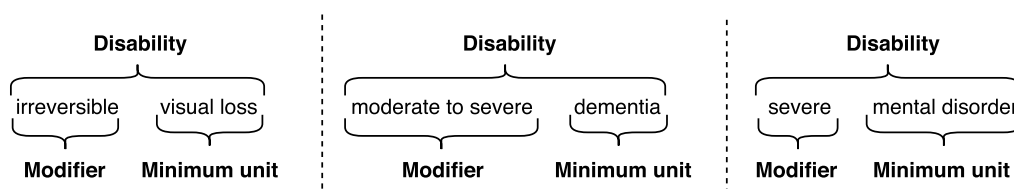


FIGURE 1. Examples extracted from the partial evaluation ground truth where the minimum unit considered for each disability is shown.

embeddings was very common, not all the supervised systems used them. UPC_3 presented runs studying embeddings from different sources (generic and specific domain), and UC3M considered both, character and sense embeddings, as well as word embeddings. IxaMed applied calculated embeddings using electronic health reports. Furthermore, amongst other NLP techniques, such as lemmatization or the extraction of suffixes/prefixes, the use of clustering and part-of-speech techniques were a popular practice, especially to reduce and to label the considered vocabulary. Part-of-speech taggers were considered by the participants to be very useful for the identification of qualifying expressions i.e. expressions which denote temporality or severity. Finally, other proposals introduced by the participants were the use of casing information and some external resources. While some teams used casing information to support the identification of abbreviations, the use of dictionaries and other external resources such as Wordnet, was mainly reported by unsupervised approaches (LSI_UNED and SINAI). Both unsupervised approaches were based on a similar pipeline architecture: candidate expression retrieval + filtering process based on external knowledge.

Finally, regarding the proposed architectures for named disability recognition, both supervised and unsupervised proposals were presented, being more frequent supervised or semi-supervised models. On the other hand, although solutions based on different kinds of algorithms (support vector machine, conditional random fields, random forest, etc.) were proposed, most of them were based on neural networks and deep learning.

C. DIANN SHARED TASK: ANALYSIS

This section aims to identify the most influential features used by the participating systems. We have also analyzed the most

frequent errors made by each system. From these studies and the results of the participant teams, we provide a number of insights for the design of systems to recognize disabilities mentioned in medical documents, that can also be useful for other kinds of entities.

In order to evaluate different aspects of the task, two types of matching criteria were applied: exact and partial. While the exact matching criterion looks for every proposed annotation to match exactly with the ground truth, the partial matching criterion looks for each disability to have at least its identified minimum unit or core contained in the ground truth. To carry out this evaluation, a file with the core of each annotation has been manually generated. Some examples included in this file can be seen in Fig. 1.

Precision, recall and f-measure are reported for each language taking into account the different matching criteria. Tables 3 and 4 show the best results obtained by each team for “named disability recognition” and “named disability recognition and negated disabilities identification, jointly”, respectively. Both tables show the results obtained for both languages using both evaluation criteria (exact and partial matching).

Concerning the recognition of disabilities, as shown in Table 3, the best systems were presented by IxaMed, UC3M and UPC_3; all of them based on Bi-LSTM and CRF. While UC3M and UPC_3 proposed a strategy based on one classifier, IxaMed opted for a cascade approach considering the annotation of disabilities and the annotation of abbreviations in different phases. LSI_UNED and IxaMed used rule systems based on casing information to detect abbreviations. This strategy showed a high performance, especially analyzing the partial evaluation results. Most systems exhibit notable differences in performance comparing partial and exact results. The system proposed by SINAI was one

TABLE 3. Results of named disability recognition using exact and partial evaluation (in brackets). The results obtained by both, supervised (S) and unsupervised (U) systems are shown.

	SPANISH			ENGLISH		
	Precision	Recall	F1	Precision	Recall	F1
IXA (S)	0.65 (0.72)	0.64 (0.71)	0.64 (0.71)	0.7 (0.76)	0.53 (0.57)	0.6 (0.65)
IXAMED (S)	0.75 (0.82)	0.81 (0.88)	0.78 (0.85)	0.78 (0.84)	0.86 (0.92)	0.82 (0.88)
UPC_3 (S)	0.81 (0.89)	0.59 (0.65)	0.68 (0.75)	0.79 (0.87)	0.6 (0.66)	0.68 (0.75)
LSI_UNED (U)	0.41 (0.84)	0.24 (0.51)	0.31 (0.63)	0.67 (0.85)	0.59 (0.76)	0.63 (0.8)
GPLSIUA (S)	0.79 (0.95)	0.17 (0.2)	0.28 (0.33)	0.88 (0.91)	0.25 (0.25)	0.39 (0.4)
SINAI (U)	0.18 (0.2)	0.41 (0.46)	0.25 (0.28)	0.22 (0.25)	0.42 (0.48)	0.29 (0.33)
UPC_2 (S)	0.73 (0.82)	0.5 (0.56)	0.59 (0.67)	0.75 (0.82)	0.56 (0.6)	0.64 (0.7)
UC3M (S)	0.8 (0.88)	0.65 (0.71)	0.71 (0.79)	0.77 (0.82)	0.72 (0.76)	0.74 (0.79)

TABLE 4. Results of the recognition of negated disabilities using exact and partial evaluation (in brackets). Results obtained by both supervised (S) and unsupervised (U) systems are shown.

	SPANISH			ENGLISH		
	Precision	Recall	F1	Precision	Recall	F1
IXA (S)	0.64 (0.7)	0.61 (0.67)	0.62 (0.69)	0.67 (0.75)	0.49 (0.55)	0.56 (0.63)
IXAMED (S)	0.74 (0.82)	0.79 (0.87)	0.77 (0.84)	0.74 (0.84)	0.81 (0.91)	0.77 (0.87)
UPC_3 (S)	0.77 (0.89)	0.55 (0.63)	0.64 (0.74)	0.77 (0.87)	0.58 (0.65)	0.66 (0.74)
LSI_UNED (U)	0.42 (0.8)	0.24 (0.46)	0.31 (0.58)	0.65 (0.84)	0.56 (0.72)	0.6 (0.78)
GPLSIUA (S)	0.65 (0.87)	0.11 (0.15)	0.2 (0.26)	0.8 (0.9)	0.23 (0.26)	0.36 (0.41)
SINAI (U)	0.15 (0.18)	0.34 (0.4)	0.21 (0.24)	0.19 (0.24)	0.39 (0.48)	0.26 (0.32)
UPC_2 (S)	0.71 (0.81)	0.48 (0.55)	0.57 (0.66)	0.72 (0.82)	0.51 (0.58)	0.6 (0.68)
UC3M (S)	0.76 (0.86)	0.56 (0.63)	0.65 (0.73)	0.74 (0.8)	0.62 (0.67)	0.68 (0.73)

of the least affected by the type of evaluation. However, this approach produced a large number of false positives, probably due to the mechanism used to filter candidate expressions according to the distance between embeddings. LSI_UNED was the other system that adopted an unsupervised approach divided into a candidate extraction phase and a heuristic-based filtering process. This system obtained competitive results, especially in English and for the partial matching criteria. The versions of Wordnet used for each language may be the reason of the performance differences between languages. In addition, LSI_UNED generated short annotations, often ignoring temporality or severity modifiers. In this respect, approaches using part-of-speech or sequence processing architectures, e.g. LSTM and CRF, had better results. LSI_UNED and SINAI accumulated some errors identifying diseases as disabilities. Finally, clustering techniques were very useful as method of generalization and representation, e.g. IXA and UPC_3, obtained very interesting results using these methods intensively. To summarize, detected errors were mostly found on the following aspects:

- Temporality and/or severity modifiers were not detected. Many of the detected errors in the exact evaluation are related to this.
- Identified diseases (e.g. Parkinson) or symptoms (e.g. headache) as disabilities. It is most observed in approaches that used external resources.

- Some disabilities described with more than 4 or 5 words were not covered, e.g. “Patient unable to perform activities of daily living autonomously...”. Both supervised and unsupervised systems have reported those errors.

Regarding the detection of negated disabilities, Table 4 shows the best results obtained by each team, considering simultaneously the detection of named disabilities and the identification of negated ones. For this evaluation, any negated disability was considered as false negative if only the disability was correctly identified and missing the trigger or the scope of the negation. Most teams used predefined lists of negation triggers and specific rules based on the concurrence of these triggers with one or more disabilities. Those systems defined the scope of a negation from the negation trigger to the last disability identified in the same sentence. UPC_2 and UPC_3 proposed solutions based on ABNER and NegEx, obtaining outstanding results.

Since the task of negation recognition was limited to negations that affected one or more disabilities, all systems filtered out the found negations based on this condition. Consequently, the results shown in Table 4 are strongly related to the ones shown in Table 3. In summary, IxaMed, UPC_3, UPC_2 and IXA reported best results in the identification of negated disabilities. Systems using ABNER or NegEx (UPC_3 and UPC_2) had better results in English than in

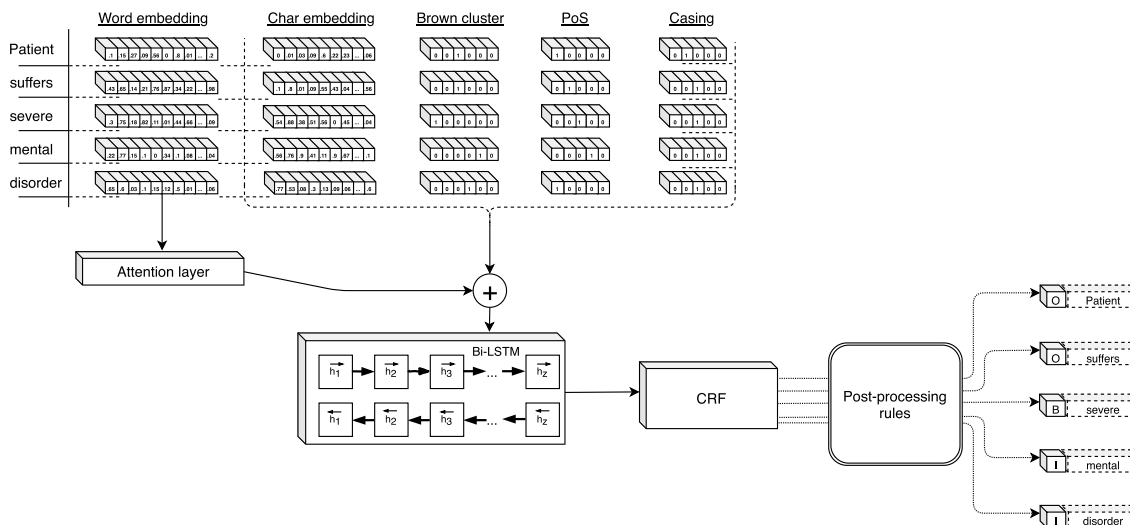


FIGURE 2. Model proposed including its inputs and layers. The post-processing rules are applied to the output of the deep learning model.

Spanish, while systems based on ad-hoc rules (IxaMed and IXA) had better results in Spanish.

D. NEW PROPOSAL

This section presents a set of experiments carried out considering some lessons learned from the presented approaches to the task on named disability recognition. We have focused the experiments on supervised approaches. Fig. 2 shows the architecture we have studied. This model is based on Bi-LSTM and CRF, in line with the teams that obtained the best results: IXAMED and UC3M teams.

We have used a Bi-LSTM and a hyperbolic tangent as activation function to process word and char embeddings, casing information, part-of-speech and Brown clusters. After analyzing different sizes for the proposed model, the size of the main Bi-LSTM is similar to the size reported by the IXAMED and UC3M teams, 150 neurons in the output layer. In addition, we have used a second Bi-LSTM with 50 neurons to process each word as a sequence of characters. The concatenation of both Bi-LSTMs is processed using a CRF. Given the small size of the corpus, we have trained this deep learning model considering small batches (16). Additionally, we have added dropouts of 0.25 between processing layers and, we have used Adam [44] as optimizer function with a learning rate of 0.01. In addition, as proposed by IXAMED, a small set of post-processing rules has been implemented. Specifically, we have implemented a set of rules to detect abbreviations (inspired by [45]) and to process special cases detected in the training set. Some of the rules are focused on the processing of enumerations, e.g.:

- If an annotation contains a statement such as: “<dis>cognitive delay/mental disability</dis>”, then it is divided into “<dis>cognitive delay</dis>” and “<dis>mental disability</dis>”.

- If an annotation contains a statement such as: “severe or <dis>moderate loss of vision</dis>”, then it is expanded to “<dis>severe or moderate loss of vision</dis>”.

The following techniques and representation methods have been studied using the described architecture:

- **Pre-trained Word embeddings (W).** The experiments were carried out using Glove [46] for both languages and working with the 200-dimensions model.
- **Char embeddings (Ch).** Due to the reduction of dimensionality caused by the use of word embeddings, some systems use this representation to reduce the loss of information. In order to obtain a sequential processing, a model based on LSTM has been used.
- **Brown cluster (B).** Cluster-based representations are very useful to support generalization capabilities of word embeddings. Brown clustering is based on the premise that expressions that occur in similar contexts could be semantically related. We modelled this feature using an one-hot vector.
- **Part-Of-Speech embeddings (P).** Part-of-speech tagging has been considered as an element to be analysed, since most of the complexity of this task consists on identifying each disability with its severity and/or temporality qualifiers.
- **Casing vector (C).** This vector encodes format information, in some cases omitted by word embeddings, e.g.: *Expression:* 01234AB – > *Casing Label:* mainly numerical expression.
Expression: Mentally – > *Casing Label:* uppercase expression.
- **Attention (A).** The attention method proposed by [47] was used to consider the filtering mechanism suggested by the SINAI team. This attention method weights each term according to its relation with certain recurrent

TABLE 5. Obtained results before applying post-processing rules using exact and partial evaluation (in brackets). The best obtained results for each language, sorting by f-measure achieved in exact evaluation are shown in boldface. (W) Word embeddings - (Ch) Char embeddings - (C) Casing - (B) Brown clusters - (A) Attention.

	SPANISH			ENGLISH		
	Precision	Recall	F1	Precision	Recall	F1
W+Ch+C	0.79 (0.90)	0.69 (0.79)	0.74 (0.84)	0.81 (0.89)	0.68 (0.75)	0.74 (0.81)
W+Ch+C+A	0.78 (0.87)	0.67 (0.75)	0.72 (0.80)	0.8 (0.86)	0.74 (0.80)	0.77 (0.83)
W+Ch+P+C+B	0.79 (0.86)	0.70 (0.76)	0.74 (0.81)	0.83 (0.88)	0.74 (0.78)	0.78 (0.83)
W+Ch+P+C+A	0.80 (0.86)	0.67 (0.73)	0.73 (0.79)	0.86 (0.92)	0.74 (0.79)	0.79 (0.85)

TABLE 6. Obtained results after applying post-processing rules using exact and partial evaluation (in brackets) and comparison with state-of-the-art systems. Best obtained results for each language, sorting by f-measure achieved in exact evaluation are shown in boldface. (W) Word embeddings - (Ch) Char embeddings - (C) Casing - (B) Brown clusters - (A) Attention.

	SPANISH			ENGLISH		
	Precision	Recall	F1	Precision	Recall	F1
IXAMED	0.75 (0.82)	0.81 (0.88)	0.78 (0.85)	0.78 (0.84)	0.86 (0.92)	0.82 (0.88)
UC3M	0.8 (0.88)	0.65 (0.71)	0.71 (0.79)	0.77 (0.82)	0.72 (0.76)	0.74 (0.79)
W+Ch+C	0.83 (0.91)	0.79 (0.87)	0.81 (0.89)	0.8 (0.90)	0.77 (0.87)	0.78 (0.89)
W+Ch+C+A	0.78 (0.86)	0.78 (0.86)	0.78 (0.86)	0.78 (0.87)	0.79 (0.88)	0.79 (0.88)
W+Ch+P+C+B	0.78 (0.85)	0.79 (0.86)	0.79 (0.85)	0.81 (0.88)	0.81 (0.89)	0.81 (0.88)
W+Ch+P+C+A	0.79 (0.85)	0.76 (0.82)	0.77 (0.84)	0.84 (0.92)	0.81 (0.89)	0.83 (0.90)

terms. Although we carried out some experiments with different sets of terms, we obtained the best results with the average of the terms “disability” and “handicap”.

IV. RESULTS

Table 5 shows the obtained results before applying post-processing rules for both languages and they were indexed according to the used features. These results show a remarkable difference in performance between exact and partial evaluation. In many cases, this difference may be explained as a result of the difficulties faced to identify contextual or modifying elements related to labeled disabilities. While in Spanish, the best results were obtained using characters and casing information, in English the best results were achieved including part-of-speech tags and the attention mechanism. Limitations of the part-of-speech model used in Spanish may be the cause of this discrepancy.

As shown in Table 6, after applying the post-processing rules we obtained notable improvements in both languages, being more evident in Spanish. On the other hand, although the obtained results show clear improvements, the high results of recall obtained by the IXAMED group stand out. These results could be a consequence of the use of specific embeddings from the biomedical domain. In summary, the proposed system improves the results obtained by the best participating system using a set of features suggested during the task.

V. DISCUSSION

During the evaluation task, two matching criteria were proposed, each focusing on a different point of the task.

Each criterion has been useful to analyze different aspects of the participant proposals. For example, since the exact matching criterion requires the identification of all possible modifiers of each identified disability, some tagging processes such as Part-of-Speech have proven to be very useful to refine the generated annotations. Considering the partial matching criterion, the analysis of the obtained results has helped us to identify elements of interest in the recognition of isolated disabilities.

Given the propensity in medical documents to use abbreviations, elements focused on the processing of this type of entities are of great interest. We have improved the identification of abbreviations considering methods such as casing or character sequence processing. The model used to represent character sequences has been quite versatile, allowing both the representation of n-grams and the representation of terms not included in the word embeddings.

On the other hand, disabilities can be expressed in countless ways. While the use of clustering techniques has supported the identification of semantically similar terms, the use of the attention model has contributed to reduce the number of false positives. However, and considering the reduced size of the corpus, an in-depth analysis of this mechanism and its effects has not been performed. Improvements using this method have only been achieved in English.

Finally, the use of post-processing rules, focusing on the improvement of the exact matching evaluation results, has been highly effective. Most of the implemented rules have been useful to deal with enumerations and abbreviations, improving the previously achieved recall results.

Concerning state-of-the-art systems, IXAMED uses both custom embeddings (generated with medical reports) and ad-hoc post-processing rules to deal with the recognition of abbreviations, which would justify the achieved results of recall.

VI. CONCLUSION

In this article, we have addressed a problem rarely addressed in the biomedical domain: the detection of named disabilities. With the support of medical doctors, we have annotated a collection of disability-related documents that can be used as a starting point for future works about collecting information about disabilities and its implications for society. Annotated documents are available both in English and in Spanish, allowing the study of multilingual approaches using the same evaluation framework. We have organized an evaluation task related to the detection of disabilities in scientific documents. This task has allowed us to study different approaches in a well-defined evaluation framework. Eight participating teams took part in the task using a wide variety of resources, techniques and approaches. Both supervised and unsupervised systems were presented.

The objective of this work has been to take advantage of an in-depth analysis of the task proposals to develop a new model that includes the most relevant aspects of each participating proposal. The new model achieved interesting improvements in both languages and the conclusions reached on each analyzed element can serve as a road-map for future research on the detection of this kind of entities, among other similar ones. Despite the good results, there is still room for improvement. We proposed the task considering the bilingualism aspect and all participants developed specific systems for each language. Differences between English and Spanish results may be due to the quality of the used external resources, especially for unsupervised systems.

Finally, the detection of negated disabilities has not been analyzed in depth by this work due to the reduced number of negations contained in the corpus. As future work, we will study the performance of different approaches in the detection of negation, using this corpus and similar ones as benchmark.

REFERENCES

- [1] W. W. Chapman, P. M. Nadkarni, L. Hirschman, L. W. D'Avolio, G. K. Savova, and O. Uzuner, "Overcoming barriers to NLP for clinical text: The role of shared tasks and the need for additional creative solutions," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 540–543, Sep. 2011.
- [2] A. Gonzalez-Agirre, M. Marimon, A. Intxaurreondo, O. Rabal, M. Villegas, and M. Krallinger, "Pharmaconer: Pharmacological substances, compounds and proteins named entity recognition track," in *Proc. 5th Workshop BioNLP Open Shared Tasks, BioNLP-OST EMNLP-IJNCLP* J. Kim, C. Nédellec, R. Bossy, and L. Deléger, Eds. Hong Kong: ACL, 2019, pp. 1–10.
- [3] I. Segura-Bedmar, P. Martínez, and M. Herrero-Zazo, "SemEval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (DDIExtraction 2013)," in *Proc. 2nd Joint Conf. Lexical Comput. Semantics, Proc. 7th Int. Workshop Semantic Eval.*, vol. 2. Atlanta, Georgia, USA: ACL, Jun. 2013, pp. 341–350.
- [4] A. Névéol, C. Grouin, X. Tannier, T. Hamon, L. Kelly, L. Goeuriot, and P. Zweigenbaum, "CLEF ehealth evaluation lab 2015 task 1b: Clinical named entity recognition," in *Proc. Work. Notes CLEF-Conf. Labs Eval. Forum*, vol. 1391, L. Cappellato, N. Ferro, G. J. F. Jones, and E. SanJuan, Eds. Toulouse, France: CEUR-WS.org, 2015, pp. 1–13.
- [5] P. Rosso, J. Gonzalo, R. Martínez, S. Montalvo, and J. Carrillo de Albornoz, Eds., in *Proc. 3rd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval) Co-Located 34th Conf. Spanish Soc. Natural Lang. Process.*, Seville, Spain, vol. 2150. CEUR-WS.org, Sep. 2018. [Online]. Available: <http://ceur-ws.org/Vol-2150>
- [6] A. R. Aronson and F.-M. Lang, "An overview of MetaMap: Historical perspective and recent advances," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 3, pp. 229–236, May 2010.
- [7] B. Tang, Y. Feng, X. Wang, Y. Wu, Y. Zhang, M. Jiang, J. Wang, and H. Xu, "A comparison of conditional random fields and structured support vector machines for chemical entity recognition in biomedical literature," *J. Cheminform.*, vol. 7, no. 1, p. 8, Dec. 2015.
- [8] B. Settles, "Biomedical named entity recognition using conditional random fields and rich feature sets," in *Proc. Int. Joint Workshop Natural Lang. Process. Biomed. Appl.*, 2004, pp. 107–110.
- [9] H. Fabregat, J. Martínez-Romo, and L. Araujo, "UNED at DIANN 2018: Unsupervised system for automatic disabilities labeling in medical scientific documents," in *Proc. 3rd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval) Co-Located 34th Conf. Spanish Soc. Natural Lang. Process.*, Seville, Spain, 2018, pp. 53–60.
- [10] M. Habibi, L. Weber, M. Neves, D. L. Wiegandt, and U. Leser, "Deep learning with word embeddings improves biomedical named entity recognition," *Bioinformatics*, vol. 33, no. 14, pp. i37–i48, Jul. 2017.
- [11] F. Li, M. Zhang, G. Fu, and D. Ji, "A neural joint model for entity and relation extraction from biomedical text," *BMC Bioinf.*, vol. 18, no. 1, pp. 198:1–198:11 Dec. 2017.
- [12] R. Weegar, A. Pérez, A. Casillas, and M. Oronoz, "Recent advances in Swedish and Spanish medical entity recognition in clinical texts using deep neural approaches," *BMC Med. Informat. Decis. Making*, vol. 19, no. 7, p. 274, Dec. 2019.
- [13] A. Casillas, A. D. de Illaraza, K. Fernandez, K. Gojenola, M. Oronoz, A. Pérez, and S. Santiso, "IXAmed-IE: On-line medical entity identification and ADR event extraction in Spanish," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, T. Tian, Q. Jiang, Y. Liu, K. Burrage, J. Song, Y. Wang, X. Hu, S. Morishita, Q. Zhu, and G. Wang, Eds. Shenzhen, China: IEEE Computer Society, Dec. 2016, pp. 846–849.
- [14] A. Yeh, A. Morgan, M. Colosimo, and L. Hirschman, "BioCreAtIvE task 1A: Gene mention finding evaluation," *BMC Bioinf.*, vol. 6, no. 1, p. S2, 2005.
- [15] M. de Chalendar, M. Daniel, A. Olry, and A. Rath, "Rare diseases and disabilities: Improving the information available with three orphanet projects," *Orphanet J. Rare Diseases*, vol. 9, no. 1, p. O31, 2014.
- [16] World Health Organization, "International classification of functioning, disability and health: Children and youth version: ICF-CY," World Health Org., Tech. Rep., 2007, p. 322.
- [17] H. Fabregat, L. Araujo, and J. Martínez-Romo, "Deep neural models for extracting entities and relationships in the new RDD corpus relating disabilities and rare diseases," *Comput. Methods Programs Biomed.*, vol. 164, pp. 121–129, Oct. 2018.
- [18] J. Reitan, J. Faret, B. Gambäck, and L. Bungum, "Negation scope detection for Twitter sentiment analysis," in *Proc. 6th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.* Lisboa, Portugal: ACL, Sep. 2015, pp. 99–108.
- [19] M. F. M. Chowdhury and A. Lavelli, "FBK-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information," in *Proc. 2nd Joint Conf. Lexical Comput. Semantics, Proc. 7th Int. Workshop Semantic Eval.*, vol. 2. Atlanta, GA, USA: ACL, Jun. 2013, pp. 351–355.
- [20] V. Vincze, G. Szarvas, R. Farkas, G. Móra, and J. Csirik, "The bioscope corpus: Biomedical texts annotated for uncertainty, negation and their scopes," *BMC Bioinf.*, vol. 9, no. 11, pp. 1–9, Dec. 2008.
- [21] R. Morante and W. Daelemans, "Conandolye-neg: Annotation of negation in Conan Doyle stories," in *Proc. 8th Int. Conf. Lang. Resour. Eval.*, Istanbul, Turkey, 2012, pp. 1563–1568.
- [22] B. Zou, Q. Zhu, and G. Zhou, "Negation and speculation identification in chinese language," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2015, pp. 656–665.

- [23] Q. Liu, F. Fancellu, and B. L. Webber, "Negpar: A parallel corpus annotated for negation," in *Proc. 11th Int. Conf. Lang. Resource Eval.*, Miyazaki, Japan, May 2018, pp. 1–9. [Online]. Available: <http://www.lrec-conf.org/proceedings/lrec2018/summaries/500.html>
- [24] S. M. Jiménez-Zafra, M. Taulé, M. T. Martín-Valdivia, L. A. Ureña-López, and M. A. Martí, "SFU ReviewSP-NEG: A Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns," *Lang. Resour. Eval.*, vol. 52, no. 2, pp. 533–569, Jun. 2018.
- [25] W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, "A simple algorithm for identifying negated findings and diseases in discharge summaries," *J. Biomed. Informat.*, vol. 34, no. 5, pp. 301–310, Oct. 2001.
- [26] N. P. Cruz, M. Taboada, and R. Mitkov, "A machine-learning approach to negation and speculation detection for sentiment analysis," *J. Assoc. Inf. Sci. Technol.*, vol. 67, no. 9, pp. 2118–2136, 2016.
- [27] F. Fancellu, A. Lopez, and B. Webber, "Neural networks for negation scope detection," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1. Berlin, Germany: ACL, 2016, pp. 495–504.
- [28] H. Fabregat, J. Martínez-Romo, and L. Araujo, "Overview of the DIANN task: Disability annotation task," in *Proc. IberEval SEPLN*, 2018, pp. 1–14.
- [29] P. López-Úbeda, M. C. Díaz-Galiano, M. T. Martín-Valdivia, and S. Jiménez-Zafra, "SINAI at DIANN-IberEval 2018. Annotating disabilities in multi-language systems with UMLS," in *Proc. 3rd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval)*, 2018, pp. 1–7.
- [30] A. R. Aronson, "Effective mapping of biomedical text to the UMLS metathesaurus: The metatmap program," in *Proc. Amer. Med. Inform. Assoc. Annu. Symp.*, Washington, DC, USA, 2001, pp. 17–21.
- [31] I. Gonaega, A. Atutxa, K. Gojenola, A. Casillas, A. D. de Ilaraza, N. Ezeiza, M. Oronoz, A. Pérez, and O. P. de Viñaspre, "A hybrid approach for automatic disability annotation," in *Proc. 3rd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval)*, 2018, pp. 31–36.
- [32] P. F. Brown, P. V. deSouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai, "Class-based n-Gram models of natural language," *J. Comput. Linguistics*, vol. 18, no. 4, pp. 467–479, 1992.
- [33] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W.-H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, Dec. 2016, Art. no. 160035.
- [34] R. Agerri and G. Rigau, "Simple language independent sequence labelling for the annotation of disabilities in medical texts," in *Proc. 3rd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval)*, 2018, pp. 25–30.
- [35] R. Agerri, J. Bermudez, and G. Rigau, "IXA pipeline: Efficient and ready to use multilingual NLP tools," in *Proc. LREC*, vol. 2014, 2014, pp. 3823–3828.
- [36] A. Clark, "Combining distributional and morphological information for part of speech induction," in *Proc. 10th Conf. Eur. Chapter Assoc. Comput. Linguistics (EACL)*, vol. 1. Stroudsburg, PA, USA: ACL, 2003, pp. 59–66.
- [37] I. Moreno, M. Romá-Ferri, and P. Moreda, "GPLSIUA team at the DIANN 2018 task," in *Proc. 3rd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval)*, 2018, pp. 1–10.
- [38] I. Moreno, M. T. Romá-Ferri, and M. P. M. Pozo, "A domain and language independent named entity classification approach based on profiles and local information," in *Proc. RANLP-Recent Adv. Natural Lang. Process. Meet Deep Learn.*, Nov. 2017, pp. 510–518.
- [39] S. Medina, J. Turmo, H. Loharja, and L. Padró, "Semi-supervised learning for disabilities detection on English and Spanish biomedical text," in *Proc. 3rd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval)*, 2018, pp. 66–73.
- [40] B. Settles, "ABNER: An open source tool for automatically tagging genes, proteins and other entity names in text," *Bioinformatics*, vol. 21, no. 14, pp. 3191–3192, Jul. 2005.
- [41] P. A. Vecino and L. Padró, "Basic CRF approach to DIANN 2018 shared task," in *Proc. 3rd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval)*, 2018, pp. 61–65.
- [42] R. M. R. Zavala, P. Martínez, and I. Segura-Bedmar, "A hybrid bi-LSTM-CRF model to disabilities named entity recognition," in *Proc. 3rd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval)*, 2018, pp. 44–52.
- [43] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2015, *arXiv:1412.6980*. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [45] S. Montalvo, M. Oronoz, H. Rodríguez Hontoria, and R. Martínez, "Biomedical abbreviation recognition and resolution by prosa-med," in *Proc. 2nd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (IberEval) Co-Located 33th Conf. Spanish Soc. Natural Lang. Process. (SEPLN)*, Murcia, Spain, 2017, pp. 247–254.
- [46] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.
- [47] A. Aker, A. Sliwa, F. Dalvi, and K. Bontcheva, "Rumour verification through recurring information and an inner-attention mechanism," *Online Social Netw. Media*, vol. 13, Sep. 2019, Art. no. 100045.



HERMENEGILDO FABREGAT was born in Alicante, Spain. He received the degree in computer science from the Polytechnic University of Valencia, in 2015, and the M.Sc. degree in computer science from the Complutense University of Madrid, in 2017. He is currently pursuing the Ph.D. degree with the Department of Lenguajes y Sistemas Informáticos (LSI), National Distance Education University (UNED), Madrid, Spain. He has been a Researcher with the LSI Department, UNED, since 2016. He has authored several articles on the application of natural language processing techniques to different kinds of documents. His main research interests include machine learning and text mining applied to the biomedical domain.



JUAN MARTÍNEZ-ROMO was born in Madrid, Spain. He received the Ph.D. degree in computer science and artificial intelligence from the Universidad Nacional de Educación a Distancia (UNED), Madrid, in 2010. He has been an Associate Professor with the Departamento de Lenguajes y Sistemas Informáticos, UNED, since 2007. He is a part of the Natural Language Processing Group, UNED. He currently holds the position of the Academic Secretary of the ETSI Informática, UNED.

He has participated in ten research projects funded in public calls, six projects with private entities, and has been a Reviewer of different indexed journals and congresses in the area. His main research interests include natural language processing, information retrieval with adversary, biomedical informatics, text mining, spam detection, web search, and use of mobile devices.



LOURDES ARAUJO received the B.S. degree in physics and the Ph.D. degree in computer science from the Universidad Complutense de Madrid, Spain, in 1987 and 1994, respectively.

After spending three years in Alcatel, she joined Universidad Complutense, in 1990. She is currently a Full Professor and the Head of the Computer Science Department, UNED (National Open University of Spain). She belongs to the Natural Language Processing and Information Retrieval Group, UNED. She has taken part, often as the Leader, in several research projects related to applications of natural language processing to different domains. She has written more than 100 research articles. Her current research interests include natural language processing and information retrieval and their application to the field of medicine.

...