Procesamiento del Lenguaje Natural, Revista nº 71, septiembre de 2023, pp. 301-311 recibido 29-06-2023 revisado 08-07-2023 aceptado 20-07-2023

MEDDOPLACE Shared Task overview: recognition, normalization and classification of locations and patient movement in clinical texts

Resumen de la tarea MEDDOPLACE: reconocimiento, normalización y clasificación de lugares y movimientos de pacientes en textos clínicos

Salvador Lima-López,[1] **Eulàlia Farré-Maduell,**[1] **Vicent Briva–Iglesias,**[2]
Luis Gasco-Sanchez,[1] **Martin Krallinger**[1]
[1]Barcelona Supercomputing Center
[2]Dublin City University
{salvador.limalopez, eulalia.farre, luis.gasco, martin.krallinger}@bsc.es,
vicent.brivaiglesias2@mail.dcu.ie

Abstract: We present the MEDDOPLACE task, the first initiative addressing the automatic detection and normalization of all location-relevant entity types present in clinical texts. The resources resulting from MEDDOPLACE can be directly useful to characterize location information of importance for disease outbreak monitoring, diagnosis and prognosis, improving patient care and safety, analyze patient movements, mobility, and travels, among many other health-related applications. MEDDOPLACE relied on a high quality manually annotated corpus of 1000 clinical cases in Spanish, together with location mention normalization (mapping to GeoNames, PlusCodes and SNOMED-CT concepts), as well as a Silver Standard dataset in multiple languages (including English, Italian, Portuguese, Dutch or Swedish). The results obtained by participating teams, as well as the generated resources show a clear practical potential to improve location analysis for health-care data processing. MEDDOPLACE resources, including detailed annotation guidelines are available at: https://zenodo.org/record/8017179.
Keywords: geoparsing, clinical departments, named entity recognition, entity linking.

Resumen: Este artículo presenta la tarea MEDDOPLACE, la primera iniciativa que aborda la detección automática y normalización de distintos tipos de localización e información relacionada (como departamentos clínicos) presentes en textos clínicos. Los recursos resultantes de MEDDOPLACE pueden ser útiles para caracterizar información sobre localizaciones importante para la monitorización de brotes de enfermedades, diagnóstico y pronóstico, mejora de la atención y seguridad del paciente, análisis de movimientos, movilidad y viajes de los pacientes, entre muchas otras aplicaciones relacionadas con la salud. MEDDOPLACE se basa en un corpus anotado manualmente de alta calidad con 1000 casos clínicos en español, estando todas sus menciones normalizadas (asociándolas a conceptos de GeoNames, PlusCodes y SNOMED-CT), así como un conjunto de datos de referencia en múltiples idiomas (incluyendo inglés, italiano, portugués, holandés o sueco). Los resultados obtenidos por los equipos participantes, así como los recursos generados, muestran un claro potencial práctico para mejorar el análisis de localizaciones en el procesamiento de datos de atención médica. Los recursos de MEDDOPLACE, incluidas las guías de anotación en castellano e inglés, están disponibles en: https://zenodo.org/record/8017179.
Palabras clave: geoparsing, departamentos clínicos, reconocimiento de entidades, normalización.

ISSN 1135-5948 DOI 10.26342/2023-71-23

1 Introduction

Location data are particularly relevant for high impact practical NLP solutions applied to different languages, content types, text genres and domains. Regarding health-related applications, information on location is crucial. Some diseases like malaria and Chagas are endemic to specific regions and geographical environments, affecting diagnostic and treatment considerations. Moreover, geolocation can be considered a risk factor in case of exposure to radiation, pollution, work-related and environmental contaminants. In an era of large immigration trends and high mobility of populations, the detection of patients' travels and movements can improve early tracking of infectious outbreaks, enabling early preventative measures. Additionally, awareness of the patient's journey through different departments/facilities within health services can be exploited to evaluate quality, safety and efficiency of care interventions, to improve planning of clinical processes, and tracing of hospital-acquired conditions. However, information on locations can usually only be found in unstructured medical documents. Doctors and other health professionals write down this essential information as free text in the medical records, underscoring the need for specialized resources aimed at location detection in the clinical domain.

Location detection is by no means a new field, especially for general domain applications using content in English. In the Message Understanding Conferences (MUC) of the 1990s, it was already defined as one of the three key types of entities to be recognized (Hirschman, 1998). The LOCATION category had a very broad scope, including proper nouns and common names of all kinds of places: cities, countries, rivers, mountains, airports, monuments, etc. This annotation system for locations prevailed throughout future influential corpora such as CoNLL 2002 (Tjong Kim Sang, 2002) and 2003 (Tjong Kim Sang and De Meulder, 2003), as well as the Spanish and Catalan resource AnCora (Taulé, Martí, and Recasens, 2008).

In 2004, the ACE (Automatic Concept Extraction) project (Doddington et al., 2004) introduced a more granular classification for entity recognition in a news corpus that included new classes and sub-classes of locations, namely: geopolitical entities (GPE; countries, cities, etc.), facilities (FAC; buildings, airports, roads, etc.) and geographic entities in general (LOC; geographic features, celestial bodies, some locative phrases, etc.). Another interesting contribution was made by the BBN (Weischedel and Brunstein, 2005) and OntoNotes (Weischedel et al., 2017) corpora, which maintained the fine-grained scheme proposed by ACE and also distinguished between proper nouns (e.g. "Spain") and generic entities (e.g. "country").

Beyond detection, the normalization and disambiguation of location entities is critical for practical exploitation scenarios. This is the objective of toponym resolution (as well as similar fields such as geoparsing or geocoding), which deal with the mapping of locations either to a set of geographic coordinates or to an identifier of an ontology like GeoNames.[1] This task is usually part of Geographic Information Retrieval (GIR) systems or methods. Some resources that stand out in this area are the multiple GeoCLEF tasks conducted between 2005 (Gey et al., 2005) and 2008 (Mandl et al., 2008), which aimed to evaluate GIR systems in news texts in different languages, or the SemEval 2019 task 12 (Weissenbacher et al., 2019), related to the recognition of toponyms in English-language PubMed scientific articles and normalization to GeoNames.

Despite this wealth of resources for general domain applications, the recognition and normalization of locations in clinical data has not been addressed yet. Some characteristics of locations in the clinical domain that showcase the need for specialized resources are (a) the prevalence of proper nouns and generic facilities, specially within healthcare and social services; and (b) the prevalence of clinical departments, a type of entity highly relevant and very specific to this domain that can sometimes refer to a location, an organization, or even a group of people. Other related information, such as movement and transportation, languages and social groups, have not been really explored from a clinical perspective yet. Finally, no Spanish corpora include a fine-grained annotation scheme for locations.

To address the current lack of location annotation and extraction resources in the

[1] `http://geonames.org/`

clinical domain, we organized the MEDDOPLACE shared task, as part of the IberLEF 2023 evaluation initiative. It was based on the MEDDOPLACE Gold Standard corpus, a collection of 1,000 clinical case reports annotated with fine-grained location mentions and domain-specific data such as clinical departments. Moreover, to enable toponym resolution and patient journey extraction, all mentions in this corpus were exhaustively normalized.

This paper presents an overview of the data, sub-tasks and results of the MEDDOPLACE shared task. It is structured as follows: Section 2 introduces the shared task, including its sub-tasks and evaluation methods. Next, Section 3 describes the MEDDOPLACE corpus and other associated resources, while Section 4 presents the participation results and proposed methodologies. Finally, Section 5 concludes the paper with a discussion and future work.

2 Task Description

2.1 Shared Task Description

MEDDOPLACE (MEDical DOcument PLAce-related Content Extraction) is a shared task about geographical information extraction/toponym resolution in the clinical domain, structured into four subtasks: (a) location and place-related entity mention detection; (b) entity normalization/geocoding to GeoNames, PlusCodes and SNOMED CT (depending on entity type); (c) location entity classification and (d) end-to-end evaluation of detection, normalization and classification. These sub-tasks are explained in Section 2.2.

The participants' predictions were evaluated using the CodaLab platform (Pavao et al., 2022), by comparing automatic predictions against manual annotations generated for the test set subset of the MEDDOPLACE corpus, relying on both classification-based metrics (precision, recall, F1-score) and distance-based metrics (Accuracy@161km, Area Under the Curve). The evaluation setting and metrics are detailed in Sections 2.3 and 2.4 respectively.

Figure 2 gives a visual overview of the shared task and its setting.

2.2 Sub-tasks

The MEDDOPLACE track was divided into the following sub-tasks (the first three focus on different steps of the information extraction pipeline, while the fourth evaluates all three previous sub-tasks' performance in succession):

- **Sub-task 1:Location Entity Recognition** Participants had to automatically detect mentions of locations and location-related entities in clinical case reports in Spanish. Using the MEDDOPLACE corpus as training data, they had to create systems able to retrieve the start and end position of the entities mentioned in the text, as well as the appropriate label.

- **Sub-task 2: Geographic Normalization** This is an entity linking/toponym resolution task. Participants were challenged to automatically normalize all mentions in the corpus. This sub-task was further divided into three independent parts: (a) normalization to GeoNames of named geopolitical and geographical entities; (b) normalization to PlusCodes of named facilities; (c) normalization to SNOMED-CT of the remaining entity types.

- **Sub-task 3: Location Entity Classification** Participants had to classify the location entities (that is, GPE, GEO and FAC entities) into four different classes of clinical relevance: (a) the patient's place of origin; (b) the patient's place of residence; (c) a place where the patient has travelled to or from; (d) a place where the patient has received medical attention. Only one label is possible for each annotation.

- **Sub-task 4: Location End-to-End Evaluation** Participant systems were evaluated in all the above three tasks sequentially, instead of being evaluated separately. They had to create systems (or a combination of systems) able to detect entities, normalize and, finally, classify them. Each step was evaluated individually and an average F1-score was provided. This type of end-to-end evaluation is more holistic and provides a more realistic assessment of the participating systems' complete performance, better reflecting their applicability and effectiveness in real-world scenarios.

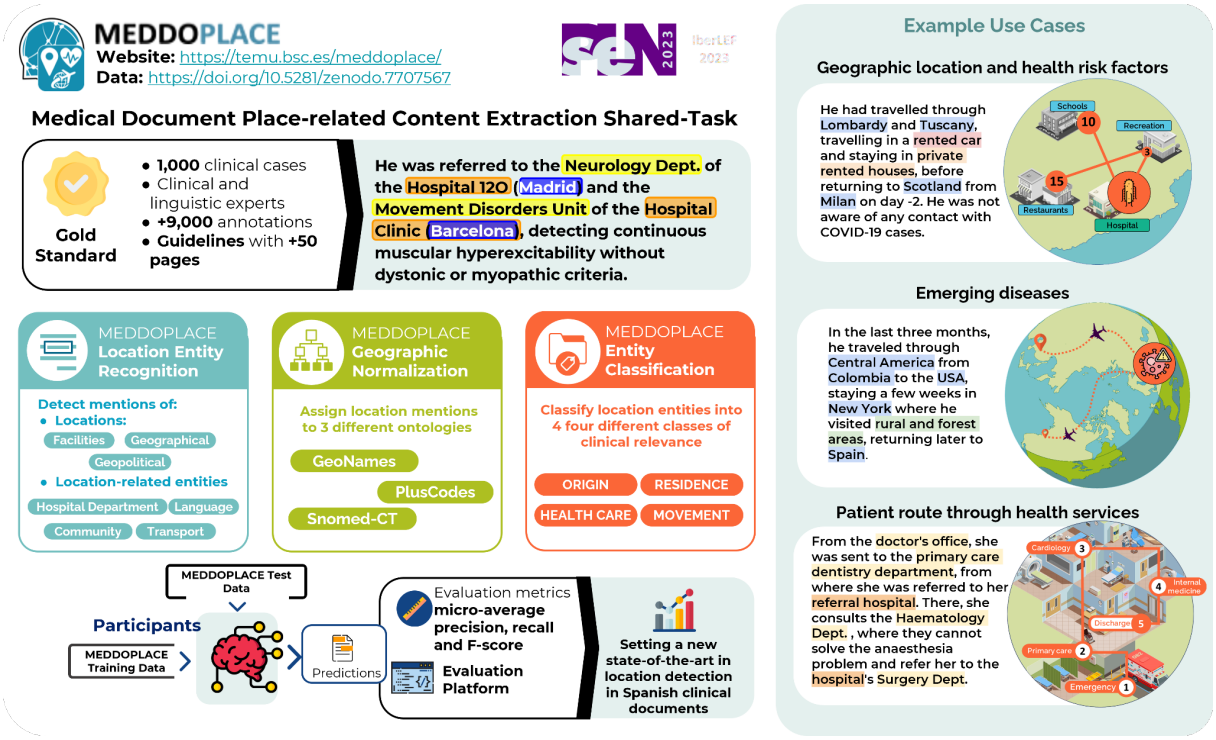

Figure 1: Visual overview of the MEDDOPLACE Shared Task. Image originally used for task dissemination.

2.3 Evaluation Setting

For the task's evaluation, we used CodaLab (Pavao et al., 2022). CodaLab is an open-source platform for conducting computational research and benchmarking in a more efficient, reproducible, and collaborative manner.

To allow the individual evaluation of normalization and classification systems, the evaluation process was separated into two phases:

- **Phase 1.** Using only the test set text files, participants had to create predictions for named entities (sub-task 1), as well as for normalization and classification based on their retrieved entities (sub-task 4, end-to-end).
- **Phase 2.** For this part, the complete list of mentions in the Gold Standard test set was released. Participants had to normalize (sub-task 2) and classify (sub-task 3) all mentions in the list.

A Task Guide[2] was released to help participants understand the evaluation setting, input and output data for every sub-task.

2.4 Evaluation Metrics

Due to the multiple sub-tasks, geographical component of the task and the evaluation setting in two phases, multiple metrics had to be used for the evaluation. They could be distinguished into two groups:

- **Classification-based metrics.** Micro-averaged precision, recall and F1-score were used to evaluate named entity recognition (sub-task 1) and end-to-end systems (sub-task 4). For the NER sub-task, both strict and overlapping versions of these metrics were provided. Additionally, accuracy (calculated as the fraction of correct mentions out of the total number of mentions) was used for the individual evaluation of normalization and classification systems.
- **Distance-based metrics.** As part of the normalization sub-task, the coding of named locations to GeoNames and PlusCodes were also evaluated using four established distance-based metrics that are actively used in toponym resolution and geocoding tasks (see for instance Gritta, Pilehvar, and Collier (2020)).

[2] https://temu.bsc.es/meddoplace/wp-content/uploads/2023/05/MEDDOPLACE-Task-Guide.pdf

Namely, these are: mean distance error, median distance error, accuracy at 161 kilometers (A@161km; the percentage of references resolved within an error distance of 161 kilometers) and area under the curve (AUC; a logarithmic calculation of the error distances that softens the effect of outlier errors)

Out of these metrics, the main ones to determine the best system are strict F1-score in Phase 1 and accuracy in Phase 2.

The MEDDOPLACE evaluation library, which is also used in the CodaLab platform, provides an easy and reproducible method to calculate all of these metrics. It is available on GitHub.[3]

3 Corpus and Resources

This section explains all the different resources released for the MEDDOPLACE shared task, including the MEDDOPLACE corpus, annotation guidelines and other additional data.

3.1 MEDDOPLACE Corpus

The MEDDOPLACE Gold Standard corpus is a collection of 1,000 clinical case reports in Spanish from various medical specialties such as psychiatry, neurology, travel medicine, infectious diseases, cardiology, occupational medicine and oncology. The corpus has been annotated, normalized and classified by a linguist with the assistance of a clinical expert. The documents used were manually selected for their relevance, and the annotations were thoroughly revised in a post-processing step. The MEDDOPLACE corpus is publicly available on Zenodo.[4]

The corpus' annotation scheme includes a total of 10 labels. On the one hand, for places and locations we define six granular labels: GPE_NOM, GPE_GEN, GEO_NOM, GEO_GEN, FAC_NOM and FAC_GEN. Two existing resources, ACE (Doddington et al., 2004) and the BBN corpus (Weischedel and Brunstein, 2005), inspire this division. From the former we take the granular division of places into three categories: geopolitical entities (GPE), facilities, constructions, and buildings in general (FAC), and natural geographical features and other types of places (LOC, which we rename to GEO here). From the latter we take the division between common names (GEN) and proper nouns (NOM), which facilitates a comprehensive annotation, and also to separate the annotation into two axes: the type of location and the type of name.

On the other hand, due to the clinical focus of the corpus, we include four more location-related labels:

- **DEPARTMENT** collects all mentions of specific sites within hospitals ("operating room"), as well as services ("emergency service", "cardiology"), units ("ICU") and other clinical elements that represent the patient's journey through the healthcare system.
- **TRANSPORT** is a fuzzy label that includes mentions related to movements of the patient (and other people), both outside and inside the health care setting. For example, "travel", "ambulance transport", "flying in an airplane", "moving by car", etc.
- **COMMUNITY** considers sociodemographic information that is often (but not always) related to a person's place of origin or residence, including ethnicities and religions.
- **LANGUAGE** includes mentions of the languages mentioned in the text, as well as linguistic problems (such as "language barrier").

Most mentions in the corpus are normalized (with a few mentions which could not be assigned a code, left as NOCODE). We used three different normalization sources due to the variety of mentions in the corpus:

- **GeoNames** is a free geographic database with more than 21 million entries. Each entry has associated metadata (such as latitude, longitude, altitude, population, etc.), as well as a unique identifier code GeoNames[5] was used for named geopolitical and geographical entities (GPE_NOM and GEO_NOM labels).
- **PlusCodes**, also known as OpenLocationCodes, is a system of codes to iden-

[3] https://github.com/TeMU-BSC/meddoplace_scoring_script

[4] https://zenodo.org/record/8017179

[5] http://geonames.org/

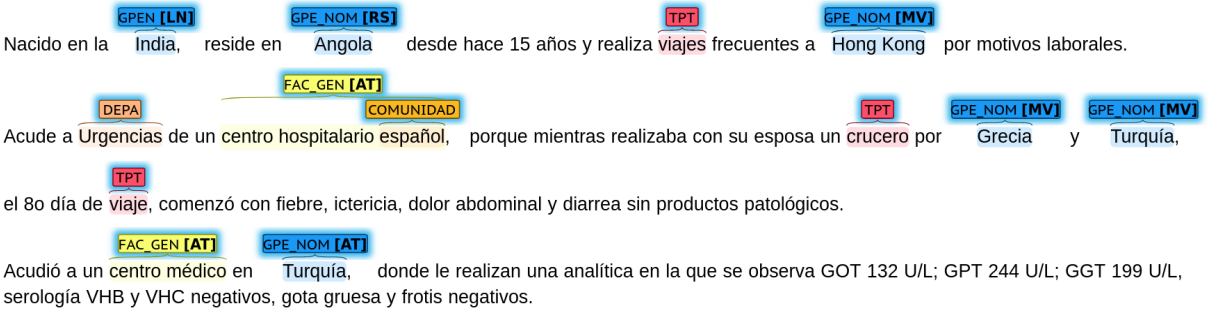

Figure 2: Example of the MEDDOPLACE corpus. Translation (marked entities in italics): "Born in *India*, they have lived in *Angola* for 15 years and *travel* frequently to *Hong Kong* for work. They visited the *emergency department* of a *Spanish hospital* because while they were on a *cruise* with their wife in *Greece* and *Turkey*, on the 8th day of the *trip*, they started with fever, jaundice, abdominal pain and diarrhea without pathological products. They visited a *medical center* in *Turkey*, where they underwent a blood test showing [...]".

tify specific locations on Earth quite accurately. To create these codes, the world is divided into small grids, each with a code. Each code consists of a series of letters and numbers, calculated from latitude and longitude, that represent a specific location on the surface of the Earth. PlusCodes[6] was used for named facilities (FAC_NOM label).

- **SNOMED CT** is a multilingual and comprehensive clinical terminology for coding multiple elements of the medical record. Among them, we find mentions of clinical services, ethnic groups and means of transportation. For this reason, SNOMED CT[7] was used for the remaining entity types (generic locations, clinical departments, communities, languages, and transports).

The final annotation layer in the corpus categorizes location entities using contextual information into four classes of clinical relevance: (a) birthplace, (b) residence, (c) movement, and (d) healthcare attention.

Figure 3 shows an example of an annotated document with some of the corpus' labels, showcasing the variety of information included in the corpus.

3.2 MEDDOPLACE Guidelines

The MEDDOPLACE guidelines describe how to annotate, normalize and classify different types of locations and related information in medical documents in Spanish. They were created *de novo* by clinical and linguistic experts after an extensive comparison and analysis of existing location corpora and their guidelines. Then, the guidelines and corpus were iteratively refined with five rounds of parallel annotation of 15 documents each. The final inter-annotator agreement (IAA) is of 0.867 for the text annotation, 0.889 for the classification and 0.857 for the normalization.

The guidelines include 67 pages and a total of 60 rules divided into different types (general, positive, negative, and special). The guidelines also explain how to normalize the annotated mentions to the three different normalization sources (GeoNames, PlusCodes and SNOMED CT), as well as how to perform the classification task.

The MEDDOPLACE guidelines are publicly available on Zenodo both in Spanish[8] and English.[9] The document was originally written in Spanish and then translated into English manually by a professional translator.

3.3 MEDDOPLACE Gazzetteer

The MEDDOPLACE gazetteer serves as an additional resource to help in the normalization of generic locations and non-location entities, such as hospital department names, to the SNOMED CT terminology. It was created using the Spanish edition of SNOMED dated October 31, 2022, where a specific subset of concepts relevant to the task was chosen. The gazetteer is meant to be used as a reference for the SNOMED normalization

[6] https://maps.google.com/pluscodes
[7] https://browser.ihtsdotools.org/
[8] https://zenodo.org/record/7775235
[9] https://zenodo.org/record/7928146

Team	Affiliation	Sub-tasks	Reference
Fade	Yunnan University, China	1	-
NLP_URJC	Universidad Rey Juan Carlos, Spain	1+3	(Roldán-Álvarez et al., 2023)
Pakapro	University of Tokyo, Japan	1+2+3+4	-
SINAI	Universidad de Jaén, Spain	1+2+3+4	(Chizhikova et al., 2023)

Table 1: Overview of the teams that participated in MEDDOPLACE. In the Affiliation column, A/I stands for academic or industry institution.

	NER		End-to-End					
Team	(Strict)	(Overlap)	NER	GN	PC	SCT	Class	Avg.
Fade	0.20	0.41	-	-	-	-	-	-
NLP_URJC	0.49	0.53	-	-	-	-	-	-
SINAI	**0.85**	**0.89**	0.85	0.66	0.26	0.69	0.66	**0.62**
Baseline	0.46	0.58	0.46	0.61	0.13	0.42	0.11	0.33

Table 2: Results of the MEDDOPLACE shared task Phase 1, which included the NER and End-to-End sub-tasks. Only F-1 scores are shown. For the End-to-End sub-task, GN stands for GeoNames, PC for PlusCodes and SCT for SNOMED CT.

sub-task.

The gazetteer comprises a collection of 23,380 concepts and a total of 27,982 lexical entries. These entries comes from various branches of SNOMED CT, including "physical object", "occupation", "environment" and "geographic location", among others. To generate the gazetteer, we used the *Gaznomed*[10] repository, along with the RF2 files of SNOMED CT. We then selected a subset of SNOMED CT branches based on the requirements of the project. Once generated, we conducted a manual review to eliminate certain obsolete lexical entries that caused ambiguity in some codes.

3.4 Additional Resources

On top of the corpus, guidelines and gazetteer, some additional resources were released as part of the task. All of these resources are also available on Zenodo.[11]

Multilingual Silver Standard. Following the methodology proposed in previous tasks like DisTEMIST (Miranda-Escalada et al., 2022), LivingNER (Miranda-Escalada et al., 2022) or MedProcNER (Lima-López et al., 2023) for the creation of multilingual resources, we have published a multilingual version of MEDDOPLACE corpus translated into English, French, Portuguese, Italian, Romanian, and Catalan. This version has been automatically created using a lexical annotation transfer approach from the Spanish Gold Standard to machine translated versions of the texts in the corpus. It is meant to be used as a starting point for the creation and adaptation of new corpora and models based on MEDDOPLACE in different languages.

Normalization cross-mapping. A version of the dataset containing MeSH descriptors obtained using the cross-mapping between SNOMED-CT and MeSH available in UMLS (Schuyler et al., 1993) has been published. This data is intended to foster the reuse of the MEDDOPLACE corpus in document retrieval models based on semantic indexing, such as those developed in the MESINESP2 task (Gasco et al., 2021; Nentidis et al., 2021). The cross-mappings correspond to Sub-task 2, which might facilitate the identification of documents containing location information.

4 Results

4.1 Participation

All in all, 20 teams registered using the official registration form and 16 in the CodaLab platform. These teams come from 12 different countries, including Canada, Turkey, Pakistan or Italy, as well as from both academic and industrial settings. Despite this, only four teams submitted their predictions for at least one of the sub-tasks. Table 1 shows a complete list of all participant teams.

4.2 Results and Methodologies

Table 2 shows the overall results for the entity recognition and end-to-end sub-tasks

[10] https://github.com/luisgasco/gaznomed

[11] https://zenodo.org/record/8017179

	GPE_NOM			GPE_GEN			GEO_NOM			GEO_GEN			FAC_NOM			FAC_GEN		
Team	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Fade	0.56	0.30	0.39	0.16	0.00	0.01	0.00	0.00	0.00	0.42	0.12	0.19	0.25	0.26	0.26	0.37	0.06	0.10
NLP_URJC	0.56	0.53	0.54	0.22	0.41	0.29	0.55	0.50	0.52	0.33	0.57	0.42	0.41	0.60	0.49	0.38	0.57	0.44
SINAI	0.90	0.88	**0.89**	0.82	0.67	**0.74**	0.71	0.50	**0.58**	0.71	0.83	**0.77**	0.68	0.75	**0.72**	0.89	0.82	**0.86**
Baseline	0.70	0.56	0.64	0.23	0.79	0.35	1.00	0.30	0.46	0.61	0.77	0.68	0.53	0.07	0.13	0.43	0.76	0.55

Table 3: Label-wise strict precision, recall and F1-score results for the location labels in the named entity recognition sub-task. The locations labels are divided in two axes: on the one hand, GPE stands for geopolitical entity, GEO for geographical location and FAC for facility; on the other hand, NOM stands for named and GEN for generic.

	Department			Transport			Community			Language		
Team	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Fade	0.26	0.14	0.18	0.45	0.08	0.14	0.19	0.14	0.16	0.00	0.00	0.00
NLP_URJC	0.47	0.69	0.56	0.48	0.44	0.46	0.59	0.20	0.30	0.00	0.00	0.00
SINAI	0.86	0.90	**0.88**	0.83	0.75	**0.79**	0.80	0.71	**0.75**	1.00	0.52	**0.68**
Baseline	0.22	0.75	0.34	0.77	0.76	0.77	0.84	0.46	0.59	0.68	0.56	0.61

Table 4: Label-wise strict precision, recall and F1-score results for the non-location labels in the named entity recognition sub-task.

(i.e. Phase 1). A more detailed, label-wise report of the named entity recognition sub-task is also shown in Tables 3 and 4. Finally, the results for Phase 2 (normalization and classification sub-tasks) are included in Table 5. Due to limited space, only each team's best run is shown. In the tables, the best result is bolded, and the second-best is underlined. A dash (-) means that the team did not participate in the sub-task. Participant team Pakapro is not included in the tables due to a submission error.

Next, we provide a short overview of the methodologies used as part of the task:

Baseline To establish a basic benchmark for participants to compare their systems to, we developed a system that employs a vocabulary transfer approach between the training and test sets. For the named entity recognition sub-task, we created a dictionary with all annotations from the training set and tried to match them to the texts in the test set. Regarding the Entity Linking sub-tasks, we assigned the same code to each mention identified in the test set as it had in the training set. Finally, for the classification sub-task, we calculate the probability distribution of the four classes in the training set and select a random class for each mention based on said distribution.

Team Fade This team developed a system for the named entity recognition task that uses a pre-trained BERT model followed by a BiLSTM unit and a GATE module. Their approach achieves a 0.20 F1-score, which is below the baseline system, partly due to some pre-processing errors.

Team NLP_URJC This team participated in two tasks: named entity recognition and entity classification. For the former, they employed a combination of RoBERTA models and BiLSTM units. For the latter, they use a pre-trained BERT model fine-tuned for sequence classification. Interestingly, they used Generative Pretrained Models (GPT) to create new sentences for some underrepresented classification classes.

Team Sinai This team participated in all sub-tasks, including the end-to-end evaluation. For named entity recognition, they tried different classifiers such as BiLSTM and CRF on top a RoBERTA model. The best score was achieved by an ensemble model that also considers nested entities. Regarding GeoNames normalization, they used training set vocabulary transfer with fuzzy matching and Levenshtein distance techniques. For PlusCodes normalization they used the search tool Nominatim[12] to extract coordinates. Next, for SNOMED CT normalization they used string matching and a SapBERT-XLMR model (Liu et al., 2021) trained with UMLS (Schuyler et al., 1993) that uses XLM-RoBERTa as a base model. Finally, their classification system uses a pre-trained RoBERTa model fine-tuned for the task. Their end-to-end participation uses the same systems described above.

[12] `https://nominatim.org/`

Team Name	GeoNames	PlusCodes	SNOMED	Classification
NLP_URJC	-	-	-	0.32
SINAI	**0.73**	**0.32**	**0.78**	**0.76**
Baseline	0.55	0.07	0.70	0.25

Table 5: Results of the MEDDOPLACE shared task Phase 2, which included the Entity Linking (divided into GeoNames, PlusCodes and SNOMED CT normalization) and Classification subtasks. Only accuracy scores are shown.

5 Discussion

We presented the MEDDOPLACE shared task and resources, a new resource for the clinical NLP community focused on the recognition, normalization, and classification of locations, clinical departments and other related information such as patient movements.

We expect that MEDDOPLACE could represent a reference resource for geoparsing not only in the clinical domain or content in Spanish due to its fine-grained annotation scheme and complete normalization. Moreover, the results of location detection systems show a multilingual adaptation potential and impact beyond healthcare. For instance, fine-grained location mentions can also be important for processing tourism/traveling-related content, legal texts, social media mining (Gasco Sánchez et al., 2022) and even noise and environmental pollution effects on health (Gasco et al., 2019).

MEDDOPLACE, along with the DisTEMIST (Miranda-Escalada et al., 2022), MEDDOCAN (Marimon et al., 2019) or MedProcNER (Lima-López et al., 2023) corpora and guidelines, contributes to an ongoing initiative aimed at fostering the creation and availability of annotated resources for extracting clinically relevant information from health-related documents. These resources have been validated by clinical and linguistic experts and are currently being used for multiple use cases, including their adaptation to hospital settings. Other resources part of this initiative include LivingNER (Miranda-Escalada et al., 2022), PharmaCoNER (Gonzalez-Agirre et al., 2019), CANTEMIST (Miranda-Escalada, Farré-Maduell, and Krallinger, 2020) and MEDDOPROF (Lima-López et al., 2021).

As future work, we expect to keep working on the geographical domain to generate and evaluate openly-available detection and normalization models, as well as apply this model to different text types and assess their relationship with other entity types such as diseases.

Acknowledgements

We acknowledge the Encargo of Plan TL (SEDIA) to BSC for funding. This project is supported by the European Union's Horizon Europe Coordination & Support Action under Grant Agreement No. 101057849 (DataTools4Heart). We also acknowledge the support from the AI4PROFHEALTH project (PID2020-119266RA-I00).

References

Chizhikova, M., J. Collado-Montañez, M. C. Díaz-Galiano, L. A. Ureña-López, and M. T. Martín-Valdivia. 2023. SINAI@MEDDOPLACE: Detecting, Normalizing, and Classifying Places and Related Information in Spanish Medical Texts. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023)*.

Doddington, G., A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel, and R. Weischedel. 2004. The automatic content extraction (ACE) program – tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, May. European Language Resources Association (ELRA).

Gasco, L., C. Clavel, C. Asensio, and G. de Arcas. 2019. Beyond sound level monitoring: Exploitation of social media to gather citizens subjective response to noise. *Science of the Total Environment*, 658:69–79.

Gasco, L., A. Nentidis, A. Krithara, D. Estrada-Zavala, R. T. Murasaki, E. Primo-Peña, C. Bojo Canales, G. Paliouras, M. Krallinger, et al. 2021. Overview of bioasq 2021-mesinesp track. evaluation of advance hierarchical

classification techniques for scientific literature, patents and clinical trials. CEUR Workshop Proceedings.

Gasco Sánchez, L., D. Estrada Zavala, E. Farré-Maduell, S. Lima-López, A. Miranda-Escalada, and M. Krallinger. 2022. The SocialDisNER shared task on detection of disease mentions in health-relevant content from social media: methods, evaluation, guidelines and corpora. In *Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task*, pages 182–189, Gyeongju, Republic of Korea, October. Association for Computational Linguistics.

Gey, F., R. Larson, M. Sanderson, H. Joho, P. Clough, and V. Petras. 2005. Geoclef: The clef 2005 cross-language geographic information retrieval track overview. volume 4022, pages 908–919, 09.

Gonzalez-Agirre, A., M. Marimon, A. Intxaurrondo, O. Rabal, M. Villegas, and M. Krallinger. 2019. Pharmaconer: Pharmacological substances, compounds and proteins named entity recognition track. In *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*, pages 1–10.

Gritta, M., M. T. Pilehvar, and N. Collier. 2020. A pragmatic guide to geoparsing evaluation: Toponyms, named entity recognition and pragmatics. *Language resources and evaluation*, 54:683–712.

Hirschman, L. 1998. The evolution of evaluation: Lessons from the message understanding conferences. *Computer Speech Language*, 12(4):281–305.

Lima-López, S., E. Farré-Maduell, L. Gascó, A. Nentidis, A. Krithara, G. Katsimpras, G. Paliouras, and M. Krallinger. 2023. Overview of medprocner task on medical procedure detection and entity linking at bioasq 2023. In *Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum*.

Lima-López, S., E. Farré-Maduell, A. Miranda-Escalada, V. Brivá-Iglesias, and M. Krallinger. 2021. Nlp applied to occupational health: Meddoprof shared task at iberlef 2021 on automatic recognition, classification and normalization of professions and occupations from medical texts. *Procesamiento del Lenguaje Natural*, 67:243–256.

Liu, F., I. Vulić, A. Korhonen, and N. Collier. 2021. Learning domain-specialised representations for cross-lingual biomedical entity linking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 565–574, Online, August. Association for Computational Linguistics.

Mandl, T., P. Carvalho, G. M. Di Nunzio, F. Gey, R. Larson, D. Santos, and C. Womser-Hacker. 2008. Geoclef 2008: The clef 2008 cross-language geographic information retrieval track overview. volume 5706, pages 808–821, 09.

Marimon, M., A. Gonzalez-Agirre, A. Intxaurrondo, M. Villegas, and M. Krallinger. 2019. Automatic de-identification of medical texts in spanish: the meddocan track, corpus, guidelines, methods and evaluation of results.

Miranda-Escalada, A., E. Farré-Maduell, S. Lima-López, D. Estrada, L. Gascó, and M. Krallinger. 2022. Mention detection, normalization & classification of species, pathogens, humans and food in clinical documents: Overview of livingner shared task and resources. *Procesamiento del Lenguaje Natural*.

Miranda-Escalada, A., E. Farré-Maduell, and M. Krallinger. 2020. 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*.

Miranda-Escalada, A., L. Gascó, S. Lima-López, E. Farré-Maduell, D. Estrada, A. Nentidis, A. Krithara, G. Katsimpras, G. Paliouras, and M. Krallinger. 2022. Overview of distemist at bioasq: Automatic detection and normalization of diseases from clinical texts: results, methods, evaluation and multilingual resources.

Nentidis, A., G. Katsimpras, E. Vandorou, A. Krithara, L. Gasco, M. Krallinger,

and G. Paliouras. 2021. Overview of bioasq 2021: the ninth bioasq challenge on large-scale biomedical semantic indexing and question answering. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 12th International Conference of the CLEF Association, CLEF 2021, Virtual Event, September 21–24, 2021, Proceedings 12*, pages 239–263. Springer.

Pavao, A., I. Guyon, A.-C. Letournel, X. Baró, H. Escalante, S. Escalera, T. Thomas, and Z. Xu. 2022. Codalab competitions: An open source platform to organize scientific challenges. *Technical report*.

Roldán-Álvarez, D., M. Rodríguez-García, M. C. Her, L. A. Ureña-López, and M. T. Martín-Valdivia. 2023. URJC-Team at MEDDOPLACE 2023: Bi-LSTM and Transformers for Medical Document Place-Related Content Extraction. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023)*.

Schuyler, P. L., W. T. Hole, M. S. Tuttle, and D. D. Sherertz. 1993. The umls metathesaurus: representing different views of biomedical concepts. *Bulletin of the Medical Library Association*, 81(2):217.

Taulé, M., M. A. Martí, and M. Recasens. 2008. AnCora: Multilevel annotated corpora for Catalan and Spanish. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).

Tjong Kim Sang, E. F. 2002. Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition. In *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*.

Tjong Kim Sang, E. F. and F. De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.

Weischedel, R. M. and A. Brunstein. 2005. BBN Pronoun Coreference and Entity Type Corpus LDC2005T33. Web Download.

Weischedel, R. M., E. H. Hovy, M. P. Marcus, and M. Palmer. 2017. Ontonotes : A large training corpus for enhanced processing.

Weissenbacher, D., A. Magge, K. O'Connor, M. Scotch, and G. Gonzalez-Hernandez. 2019. SemEval-2019 task 12: Toponym resolution in scientific papers. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 907–916, Minneapolis, Minnesota, USA, June. Association for Computational Linguistics.