

Overview of IberLEF 2023: Natural Language Processing Challenges for Spanish and other Iberian Languages

Salud María Jiménez-Zafra¹, Francisco Rangel² and Manuel Montes-y-Gómez³

¹*SINAI, Computer Science Department, CEATIC, Universidad de Jaén, Jaén, Spain*

²*Symanto Research, Valencia, Spain*

³*Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico*

Abstract

IberLEF is a shared evaluation campaign of Natural Language Processing systems in Spanish and other Iberian languages that has been organized since 2019, and is held as part of the annual conference of the Spanish Society for Natural Language Processing. Its goal is to encourage the research community to organize competitive text processing, understanding and generation tasks in order to define new research challenges and set new state-of-the-art results in at least one of the following Iberian languages: Spanish, Portuguese, Catalan, Basque or Galician. This paper summarizes the evaluation activities carried out in IberLEF 2023, which included 14 tasks and 34 subtasks dealing with automatically text generated identification, clinical content, code switch analysis, early risk prediction on the Internet, harmful and inclusive content detection, political ideology and propaganda identification, and sentiment, stance and opinion analysis. Overall, the IberLEF activities represented a remarkable collective effort involving 432 researchers from 35 countries in Europe, Asia, Africa, Australia and the Americas.

Keywords

Natural Language Processing, Artificial Intelligence, Evaluation, Evaluation Challenges

1. Introduction

IberLEF is a shared evaluation campaign of Natural Language Processing systems in Spanish and other Iberian languages that has been organized since 2019, and is held as part of the annual conference of the Spanish Society for Natural Language Processing (SEPLN). It aims to promote research in text processing, understanding and generation tasks in at least one of the following Iberian languages: Spanish, Portuguese, Catalan, Basque or Galician.

In this shared evaluation campaign, the research community defines new research challenges and proposes tasks to advance the state of the art in Natural Language Processing (NLP). These tasks are reviewed by the members of the steering and program committees of IberLEF, and

IberLEF 2023, September 2023, Jaén, Spain

✉ sjzafra@ujaen.es (S. M. Jiménez-Zafra); kico.rangel@gmail.com (F. Rangel); mmontesg@inaoep.mx (M. Montes-y-Gómez)

🌐 <https://sjzafra.github.io/> (S. M. Jiménez-Zafra); <https://www.linkedin.com/in/kicorangel/> (F. Rangel); <https://ccc.inaoep.mx/~mmontesg/> (M. Montes-y-Gómez)

🆔 0000-0003-3274-8825 (S. M. Jiménez-Zafra); 0000-0002-6583-3682 (F. Rangel); 0000-0002-7601-501X (M. Montes-y-Gómez)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

finally evaluated by the IberLEF general chairs. The organizers of the accepted tasks, set up the evaluation according to the proposal submitted, promote the task and manage the submission and scientific evaluation of the system description papers submitted by the participants. These scientific papers are included in this IberLEF proceedings volume published at CEUR-WS.org. Moreover, the task organizers have to prepare and submit an overview of their task evaluation exercise. These overviews are reviewed by the IberLEF organizing committee and then published in the journal *Procesamiento del Lenguaje Natural*, vol. 71 (September 2023 issue). Finally, the task organizers report the results of the tasks and the selected participants present the description of their systems at the IberLEF workshop.

IberLEF 2023 is held on September 26, 2023 in Jaén (Andalusia, Spain), within the framework of the XXXIX International Conference of the Spanish Society for Natural Language Processing (SEPLN 2023). This year 14 shared tasks have been accepted for IberLEF 2023, out of a total of 22 proposals. They are NLP tasks on automatically text-generated identification, clinical content, code switch analysis, early risk prediction on the Internet, harmful and inclusive content detection, political ideology and propaganda identification, and sentiment, stance and opinion analysis.

In this paper we summarize the tasks organized in IberLEF 2023, analyzing them for a better understanding of this collective effort.

2. IberLEF 2023 Tasks

The 14 tasks involved in IberLEF 2023 are presented below, grouped by theme.

2.1. Automatically Generated Texts Identification

AuTextification [1] is a multi-domain machine-generated text detection and attribution task. It consists of two subtasks: i) Subtask 1, Machine-generated text detection and ii) Subtask 2, Machine-generated text attribution. Subtask 1 is a binary classification task in which, given and Spanish or English text, it should be determined if the text has been automatically generated or by contrast it is from a human. Subtask 2 is a multi-class classification task which consists of, given an automatically generated text in Spanish or English, identifying which text model has generated it. The possible classes are A, B, C, D, E or F. Each class represent a text generation model and the model label mapping is: "A" - "bloom-1b7", "B" - "bloom-3b", "C" - "bloom-7b1", "D" - "babbage", "E" - "curie", "F" - "text-davinci-003". The domains covered in this task are tweets, reviews, news, legal, and how-to articles.

2.2. Clinical Content

ClinAIS [2] shared task aims to tackle the identification of seven section types within unstructured clinical records in the Spanish language. Specifically, the sections types are: i) Present Illness, ii) Past Medical History/ Medical History, iii) Family History, iv) Exploration, v) Evolution, vi) Treatment and, vii) Derived from/to. The dataset used in this task is a subset of the CodiEsp corpus [3], a collection of Spanish unstructured clinical case reports from different medical specialties which was used in a Named Entity Recognition task at CLEF eHealth 2020.

MEDDOPLACE [4] is a shared task about geographical information extraction and toponym resolution in the clinical domain. It is structured into four subtasks: i) Location and place-related entity mention detection, ii) Entity normalization (geocoding to GeoNames, PlusCodes and SNOMED CT), iii) Location entity classification and, iv) End-to-end evaluation of detection, normalization and classification. The corpus of this task consists of 1,000 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).

TESTLINK [5] focuses on relation extraction from clinical cases in Spanish and Basque. It consists in identifying textual mentions of both laboratory tests and their results in a clinical narrative, and then linking tests to their respective results. The task is divided into two tracks depending on the language: i) Spanish and ii) Basque. The dataset used is based on the Spanish and Basque parts of E3C, the multilingual European Clinical Case Corpus [6], which consists of three sections of clinical cases published in medical journals and other medical resources.

2.3. Code Switch Analysis

GUA-SPA [7] is a shared task for detecting and analyzing code-switching in Guarani and Spanish. This challenge consists of three subtasks: i) Language identification in code-switched data, i.e., identifying the language of each token of a given text; ii) Named entity classification; and, iii) Spanish code classification, which consists of classifying the way a Spanish span is used in the code-switched context. The corpus of this task consists of 1,500 texts extracted from news articles and tweets, around 25 thousand tokens.

2.4. Early Risk Prediction on the Internet

MentalRiskES [8] aims to promote the early detection of mental risk disorders in Spanish. This task must be resolved as an online problem, that is, the participants must be able to detect a potential risk as early as possible in a continuous stream of data. It includes three subtasks: i) Eating disorders detection, ii) Depression detection and, iii) Non-defined disorder detection on an undisclosed disorder during the competition (anxiety) to observe the transfer of knowledge among the different disorders proposed. Participants were also asked to submit measurements of carbon emissions for their systems, emphasizing the need for sustainable NLP practices. For this competition, a set of comments from Telegram users was compiled.

2.5. Harmful and Inclusive Content

DA-VINCIS [9] supports research into the development of automatic solutions for detecting violent events in social networks. It proposes two subtasks: i) A binary classification task aimed to determine whether a tweet is about a violent incident or not and, ii) A multi-label multi-class classification task in which the category(ies) of a violent incident must be identified. This shared task was also organized in IberLEF 2022 [10]. In this edition, instead of only providing textual data the participants were provided with a multimodal dataset consisting of Mexican Spanish tweets associated with at least an image.

HOMO-MEX [11] encourages the development of NLP systems for detecting and classifying LGBTQ+ phobic content in Mexican Spanish tweets. This shared task is divided into two tracks: i) Determining whether a tweet exhibit LGBTQ+ phobic content or not and, ii) Classifying the LGBTQ+ phobic tweets as containing Lesbophobia (L), Gayphobia (G), Biphobia (B), Transphobia (T), and/or other LGBTQ+phobia (O).

HOPE [12] shared task is related to the inclusion of vulnerable groups and focuses on the study of the detection of hope speech, in pursuit of Equality, Diversity and Inclusion (EDI). It consists of, given a text, written in Spanish or English, identifying whether it contains hope speech or not. It is divided into two subtasks, according to the language in which the texts are written: i) Identifying whether a Spanish tweet contains hope speech or not and, ii) Determining whether an English YouTube comment contains hope speech or not.

HUHU [13] focus is on examining the use of humour to express prejudice towards minorities, specifically analyzing Spanish tweets that are prejudicial towards: women and feminists, LGBTIQ community, immigrants and racially discriminated people, and overweight people. This shared task consists of three subtasks: i) Determining whether a prejudicial tweet is intended to cause humour, ii) Identifying the targeted groups (women and feminists, LGBTIQ community, immigrants and racially discriminated people, and overweight people) on each tweet as a multilabel classification task and, iii) Predicting how prejudicial a message is on average to minority groups on a continuous scale from 1 to 5.

2.6. Political Ideology and Propaganda

DIPROMATS [14] is organized with the aim of finding the best techniques to identify and categorize propagandistic tweets from governmental and diplomatic sources. It presents three subtasks for each language, Spanish and English: i) A binary classification task to decide whether a tweet contains propaganda techniques, ii) A multiclass, multilabel classification task, where systems have to decide, for each tweet, in which of the 5 available categories it fits and, iii) A fine-grained classification task in which systems have to decide which of the available techniques the tweet contained.

PoliticES [15] goal is to extract political ideology and other psychographic and demographic characteristics of users in social networks. For this purpose, a cluster profiling task is proposed. It focuses on the identification of two demographic traits (self-assigned gender and profession) and one psychographic trait (political ideology), from a binary and multi-class perspective, from clusters of Spanish tweets posted by users who share these traits. This shared task was also organized in IberLEF 2022 [16]. The novelty this year is that instead of profiling users, participants work with clusters of texts written by different users but with the same characteristics, in order to avoid legal and ethical issues. In addition, users who are celebrities have also been included.

2.7. Sentiment, Stance and Opinions

FinancES [17] shared task aims to extend the challenge of sentiment analysis in Spanish to the financial domain, in order to extract the sentiment that a piece of financial information can have for several actors, including the main economic target (i.e., the specific company or asset where

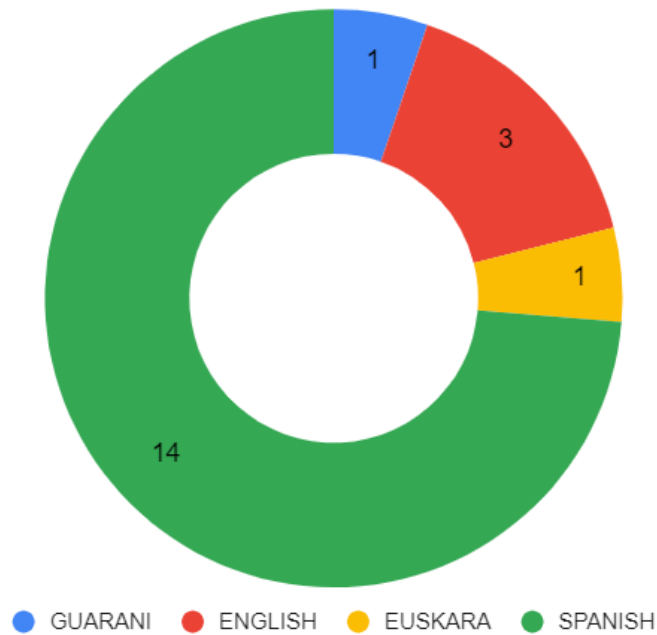


Figure 1: Distribution of languages in IberLEF 2023 tasks.

the economic fact applies), other companies (i.e., the entities producing the goods and services that others consume) and consumers (i.e., households/individuals). It consists of two subtasks: i) Identifying the main economic target from financial news headlines and determining the sentiment polarity (positive, neutral or negative) towards such target and, ii) Determining the sentiment polarity of each news headline towards both companies and consumers.

Rest-Mex [18] focuses on sentiment analysis and text clustering of tourist texts. It is divided into two subtasks: i) Sentiment analysis, to predict the polarity of opinions expressed by tourists, classifying the type of place visited (tourist attraction, hotel, or restaurant), as well as the country it is located in (Mexico, Cuba or Colombia) and, ii) Text clustering, to classify news articles related to tourism in Mexico by topic. This shared task was also organized in IberLEF 2022 [19] focusing on texts related to tourist destinations in Mexico, but this edition includes data from Cuba and Colombia for the first time.

3. Aggregated Analysis of IberLEF 2023 Tasks

3.1. Tasks characterization

In terms of **languages**, the distribution per tasks (including subtasks) is shown in Figure 1. One more year, Spanish is the central language of IberLEF (14 tasks) followed by English in a secondary role (3 tasks), and both Euskara and Guarani in a third position (1 task each). Main Spanish variants considered are those from Spain and Mexico, albeit this year for the first time

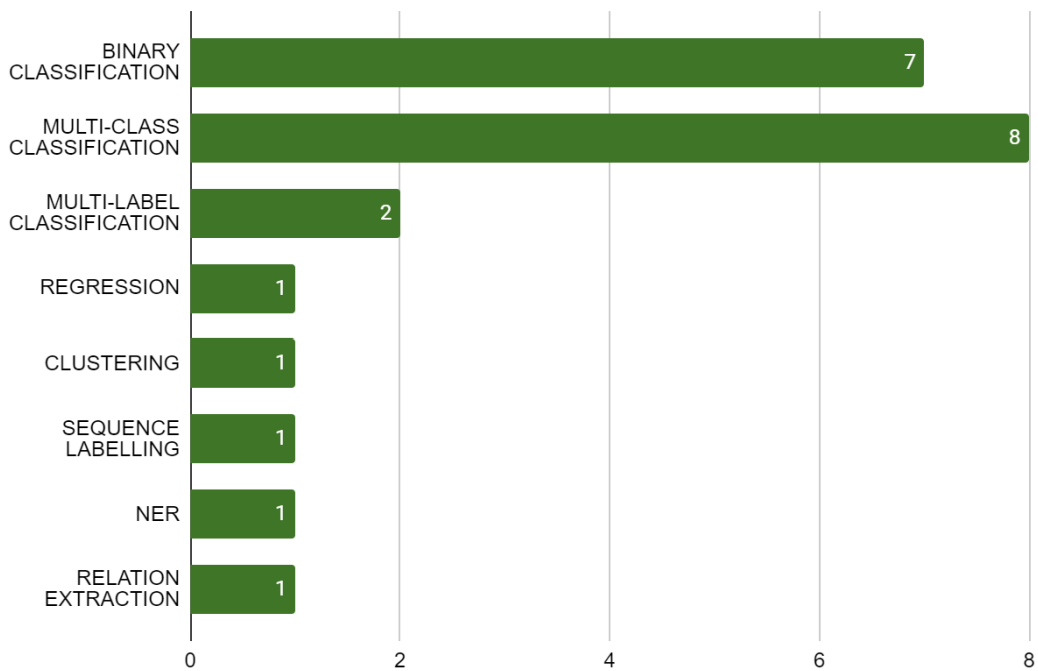


Figure 2: Distribution of IberLEF 2023 tasks per abstract task type.

Cuban or Colombian texts have been considered in one task.

In terms of **abstract task types**, the distribution of different identified subtasks can be seen in Figure 2. The most popular type of task is multi-class classification (8 tasks), followed by binary classification (7 tasks). There are also two multi-label classification tasks and one task for each regression, clustering, sequence labelling, NER and relation extraction.

Even though following the trend towards a less numerous but more diverse and complex set of tasks that started last year, this year again binary classification has been almost the most popular type of task.

In terms of **evaluation metrics**, the distribution can be seen in Figure 3, which depicts only the main metrics used to rank systems in each task. As in previous years, there is a remarkable predominance of F1 (12 tasks, in five cases together with Precision and Recall). Accuracy is used in three tasks, while MPA, CFD and RMSE in two more tasks. There are other eighteen metrics used in six tasks where each of these metrics was used only in one task. Among others, weighted B2, BPA, ICM, AUC, ERDE_x (with different values for the *x*), Pearson, P@*x* (for different *x* values). Some of them correspond to the complex tasks which embed subtasks (e.g., *Res_P*, *Res_T*, *Res_C*, Sent, Thematic, or Easiness). In others, they correspond to different distance metrics (e.g., Mean Distance Error, Median Distance Error, or A@161km).

Overall, in IberLEF as in other NLP competitive evaluation challenges we might still be relying too much on averages to combine different quality metrics: it has been common this year to combine F1 measures (which are harmonic averages) with other measures using some

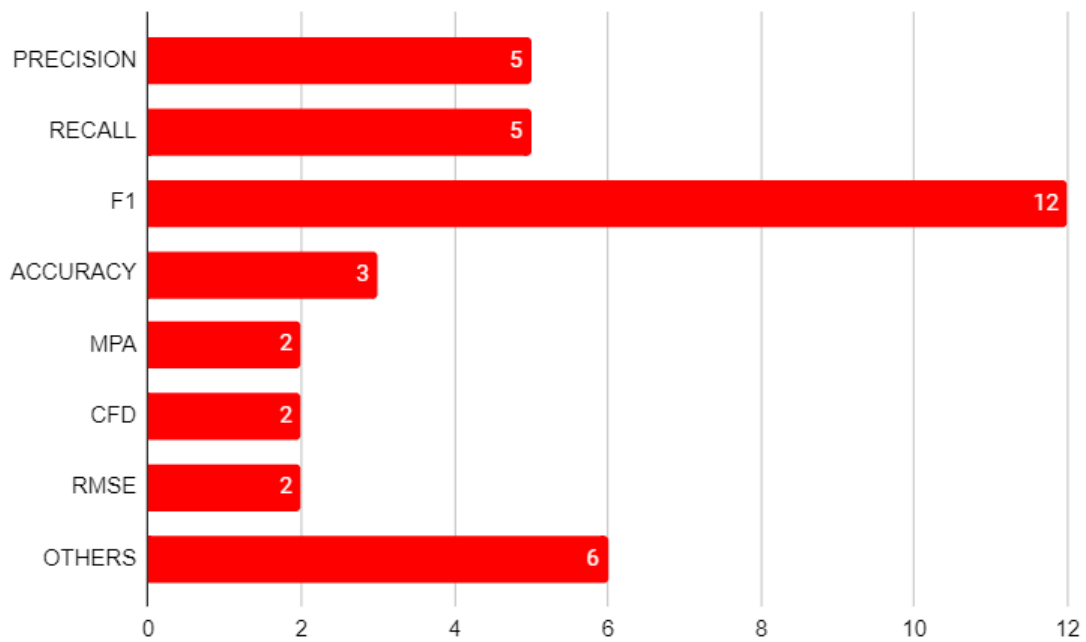


Figure 3: Distribution of official evaluation metrics in IberLEF 2023 tasks.

other form of averaging. This hides the actual behaviour of systems and give usually no clues on how to improve them. Also, again in 2023 the choice of metrics is, in general, barely justified, particularly in terms of how the system output is going to be used in realistic usage scenarios.

Finally, in terms of **novelty/stability** IberLEF 2023 has brought many new problems, with eleven out of the fourteen primary tasks being new this year (79%). Only DA-VINCIS, PoliticES, and Rest-Mex had also been run in 2022.

3.2. Datasets and results

In terms of **types of textual sources**, Figure 4 shows how they are used in IberLEF 2023 tasks. Like in 2022, there is more diversity than previous years, with new sources like Youtube, Telegram or How-to articles being considered this year. However, Twitter has become the dominant source again with more than half of the tasks (8 out of 14) using it. News and Clinical Cases Reports being used in three tasks are followed by Reviews (in one case coming from TripAdvisor) used in two tasks. The new sources (Youtube, Telegram and How-to Articles) have been used in one task each.

In terms of **dataset sizes** and annotation efforts, it is difficult to establish fair comparisons, because of the diversity of data sources, text sizes and the wide variance in terms of annotation difficulty.

In any case, in the majority of cases (11 tasks), datasets have been manually annotated. In 9 of these cases, the size of the dataset is below 10,000 instances while in two of the cases, the size is

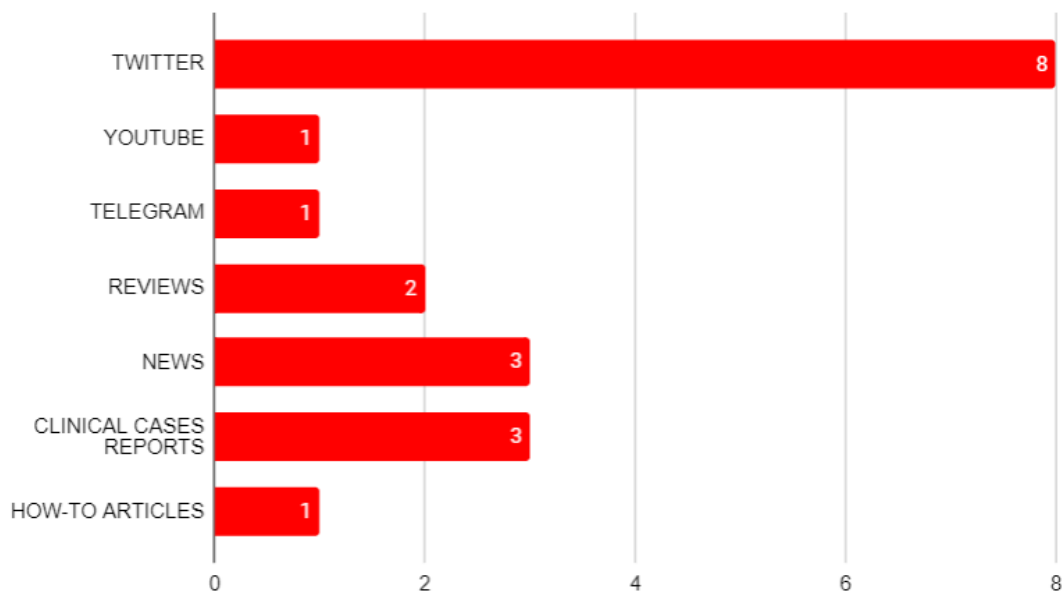


Figure 4: Types of textual sources in IberLEF 2023 tasks.

around 11,000 (HOMO-MEX) and 32,000 (HOPE) respectively. Two tasks provide self-annotated datasets with around 2,700 (PoliticES) and 360,000 (Rest-Mex) instances respectively. In the case of AuTextification, the dataset has been partially self-annotated and partially human-assisted auto-generated, providing around 160,000 instances.

As for the reliability of the annotations, one useful indicator is inter-annotator agreement, which is reported in 6 out of 14 tasks. In the tasks where it is reported, annotator agreement is high in four cases, mid-high in one, and mid-low in another one. In general, mid-low agreement may indicate the complexity of the task rather than poor annotation guidelines.

Overall the annotation effort in IberLEF 2023 keeps being a remarkable contribution to enlarge test collections for Spanish (and, less prominently, other languages). One more year, IberLEF has been carried out without specific funding sources (other than those obtained individually by the teams organizing and participating in the tasks). A centralized funding schema could certainly help reach larger and better annotations in IberLEF as a whole.

In terms of **progress with respect the state of the art**, it is as usual difficult to extract aggregated conclusions for the whole IberLEF effort, in particular given the diversity of approaches for providing task baselines: for instance, no baseline was provided in two tasks. Furthermore, in eleven subtasks, only a trivial baseline was included in the comparisons (e.g. majority class or random baselines in classification). Ten subtasks used BOW kind of approaches as baseline, while eleven used some variant of transformers (BETO, DEBERTA, ROBERTA, etc.). In some of the tasks, several baselines have been given and compared to the participants' approaches, generally combining majority, BOW, and transformers.

In the subtasks that used baselines, the baseline was beaten (by a margin larger than 5%) by the best system in 27 cases, while in 6 cases the baseline obtained better results. Looking at

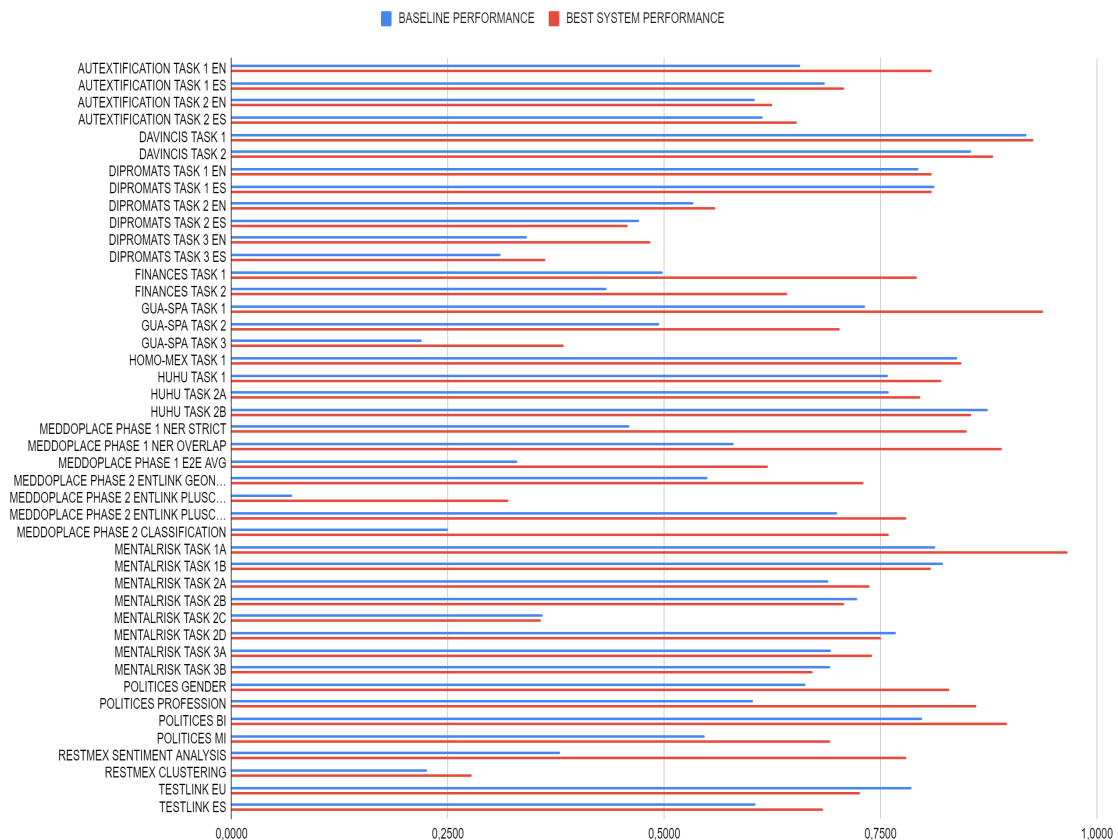


Figure 5: Performance of best systems versus baselines in IberLEF 2023 tasks. Only tasks with official evaluation metrics in the range [0-1] that include at least a baseline system are included in this graph. In the case of RMSE, it has been normalized to [1-RMSE].

the obtained results, there are only three subtasks where the best performing system obtained results higher than 0.9, and only one subtask where the baseline did. This is an indication that, in at least some of the tasks, there is still room for improvement.

In Figure 5 we display a pairwise comparison between the best system and the best baseline, for each of the tasks where at least one baseline is provided, and with respect to the official ranking metric used in each task. To avoid confusion, we have restricted the chart to tasks where the official metric varies between 0 (worst quality) and 1 (perfect output). And in the case of RMSE, we have normalized the value to [1-RMSE].

3.3. Participation

Given that IberLEF 2023 was not a funded initiative, participation has again been impressive, with a large fraction of the current research groups interested in NLP for Spanish and other Iberian languages organizing and/or participating in one or more tasks. Overall, 432 researchers representing 211 research groups from 35 countries in Europe, Asia, Africa, Australia and

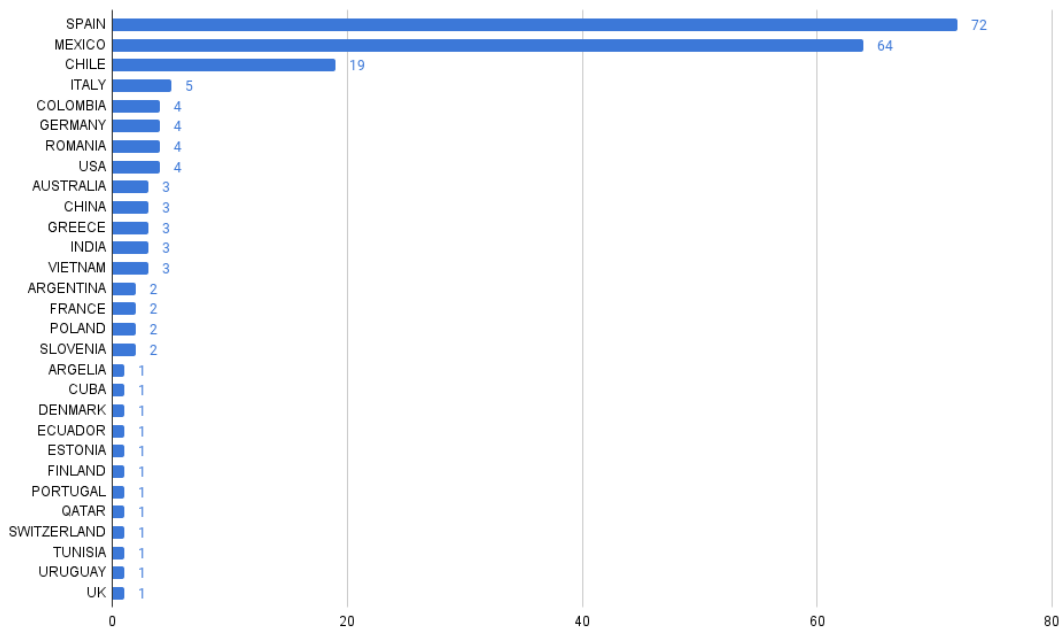


Figure 6: Number of research groups participating in IberLEF 2023 tasks per country.

the Americas were involved in IberLEF tasks. *NOTE:* Statistics have been compiled from the submitted working notes, meaning two things: *i)* Some groups and researchers may be counted twice if they have participated in more than one task; *ii)* Real participation may be higher due to the number of teams who submitted runs but did not submit their working notes afterwards and thus have not been counted in the statistics.

Figure 6 shows the distribution of research groups per country. This year, Spain has the largest representation, with 72 groups, followed by Mexico with 64 groups.

Figure 7 shows the distribution of researchers (appearing as authors in the working notes) per country. The top five, with Spain, Mexico, Chile, Colombia, and USA, represents roughly 80% of the researchers involved. In addition, it can be observed a great diversity of non-Spanish speaking countries, such as USA, Australia, India, Romania, China and Italy in the top ten positions in terms of participation, which indicates: first, that Spanish attracts the attention of the NLP community at large; and second, that current NLP technologies enable addressing different languages without language-specific machinery, other than pre-trained language models made available to the research community.

Figure 8 shows the number of teams participating in each of the tasks, considering that they submitted at least one run. Participation ranges between 3 and 46 teams. The distribution of research groups per task is shown in Figure 9. In this case, participation ranges between 3 and 43 groups. Regarding the most participated ones, except probably for the case of AuTextification, there does not seem to be a correlation between the number of participating teams and the number of participant groups. For instance, in Rest-Mex 43 groups collaborated to participate

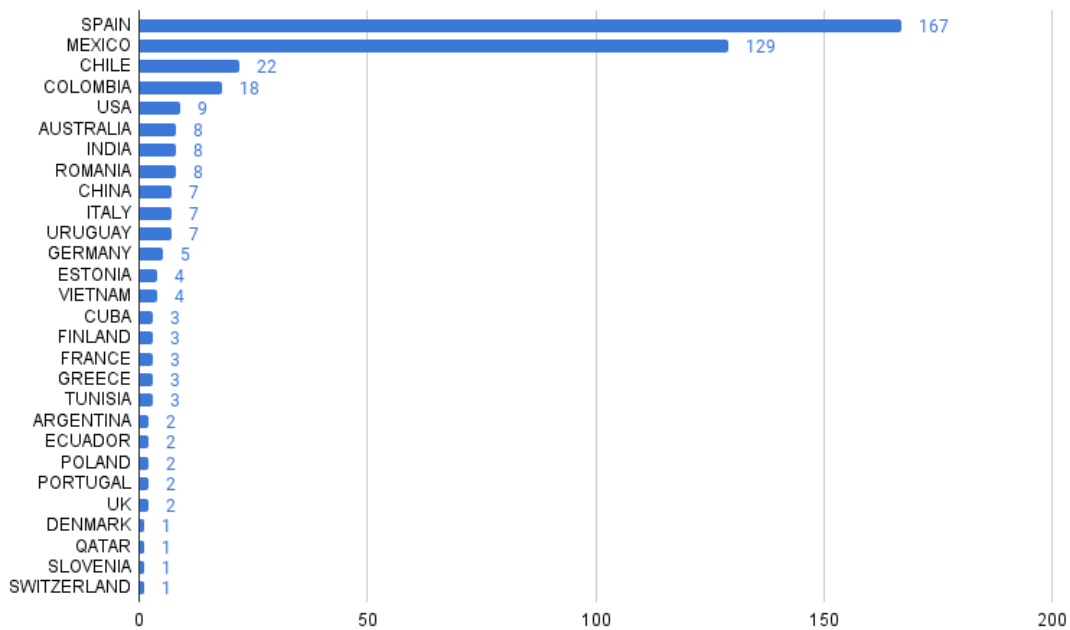


Figure 7: Number of researchers participating in IberLEF 2023 tasks per country.

as 16 teams, while in HUUH only 12 groups participated as 46 teams. *NOTE:* A team is a group of researchers coming from the same or different research groups and/or research entities who join efforts to participate in a shared task. A research group is a group of researchers often from the same faculty, specialised in the same subject, working together on the issue or topic in an official manner and not just for participating in a shared task.

As in other evaluation initiatives, participation seems to be driven not only by the task’s intrinsic interest, but also by the cost of entry: as usual, classification tasks (the most basic machine learning task, for which more plug-and-play software packages exist) receive more participation than tasks which require more elaborated approaches and more creativity to assemble algorithmic solutions.

4. Conclusions

In its fifth edition, IberLEF has again been a remarkable collective effort for the advancement of Natural Language Processing in Spanish and other Iberian languages, comprising 14 main tasks and 432 researchers involved, from institutions in 35 countries in Europe, Asia, Africa, Australia and the Americas. In comparison to the last edition, there has been a great increase in participants (28% increase, from 310 to 432) and countries (31% increase, 24 to 35), showing the increasing interest that IberLEF arouses throughout the world.

IberLEF 2023 has been one of the most diverse in terms of types of tasks and application domains, and has contributed to advance the field in the areas of automatically text generated

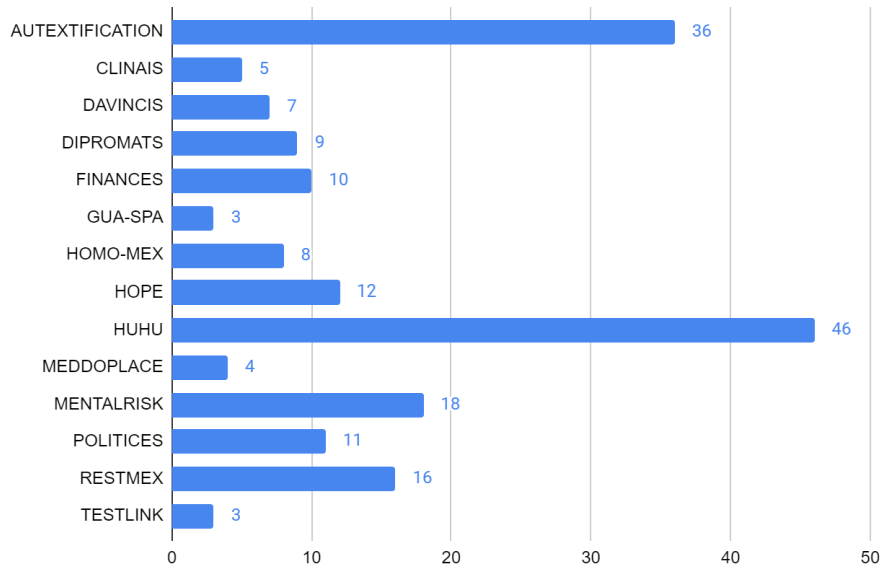


Figure 8: Distribution of participating teams per task in IberLEF 2023. The figure displays the number of teams that submitted at least one run.

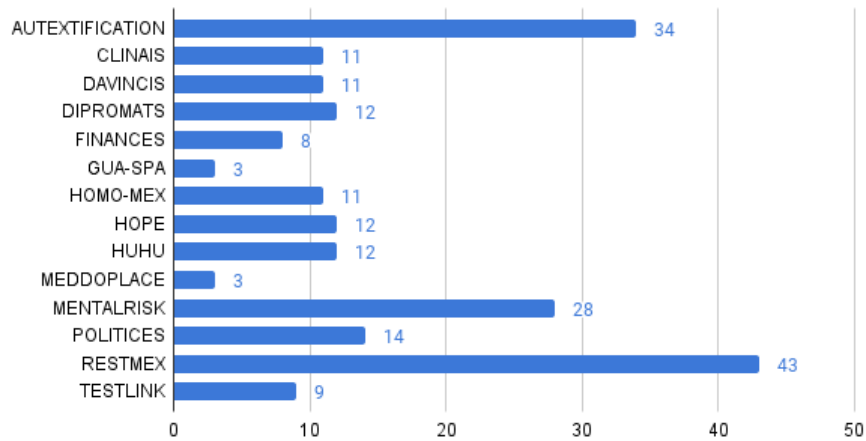


Figure 9: Distribution of participant groups per task in IberLEF 2023. The figure displays the number of groups that submitted at least one run.

identification, clinical content, code switch analysis, early risk prediction on the Internet, harmful and inclusive content detection, political ideology and propaganda identification, and sentiment, stance and opinion analysis.

In a field where machine learning is the ubiquitous approach to solve challenges, the definition of research challenges, the development of high-quality test collections that allow for iterative evaluation and the design of sound evaluation methodologies and metrics are perhaps the most critical aspects of research, and we believe IberLEF keeps making significant contributions to

all of them.

Acknowledgments

The work of the first author has been partially supported by Project CONSENSO (PID2021-122263OB-C21), Project MODERATES (TED2021-130145B-I00) and Project SocialTox (PDC2022-133146-C21) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, Project PRECOM (SUBV-00016) funded by the Ministry of Consumer Affairs of the Spanish Government, Project FedDAP (PID2020-116118GA-I00) and Project Trust-ReDaS (PID2020-119478GB-I00) supported by MICINN/AEI/10.13039/501100011033, and WeLee project (1380939, FEDER Andalucía 2014-2020) funded by the Andalusian Regional Government. Salud María Jiménez-Zafra has been partially supported by a grant from Fondo Social Europeo and the Administration of the Junta de Andalucía (DOC_01073). The work of the second author has been partially funded by the Pro²Haters - Proactive Profiling of Hate Speech Spreaders (CDTi IDI-20210776), the XAI-DisInfodemics: eXplainable AI for disinformation and conspiracy detection during infodemics (MICIN PLEC2021-007681), the OBULEX - *OBservatorio del Uso de Lenguaje sEXista en la red* (IVACE IMINOD/2022/106), and the ANDHI - ANomalous Diffusion of Harmful Information (CPP2021-008994) R&D grants.

References

- [1] A. M. Sarvazyan, J. González, M. Franco-Salvador, F. Rangel, B. Chulvi, P. Rosso, Overview of AuTextification at IberLEF 2023: Detection and Attribution of Machine-Generated Text in Multiple Domains, *Procesamiento del Lenguaje Natural 71* (2023).
- [2] I. De la Iglesia, M. Vivó, P. Chocrón, G. de Maeztu, K. Gojenola, A. Atutxa, Overview of ClinAIS at IberLEF 2023: Automatic Identification of Sections in Clinical Documents in Spanish, *Procesamiento del Lenguaje Natural 71* (2023).
- [3] A. Miranda-Escalada, A. Gonzalez-Agirre, J. Armengol-Estapé, M. Krallinger, Overview of automatic clinical coding: Annotations, guidelines, and solutions for non-english clinical cases at codiesp track of clef ehealth 2020., *CLEF (Working Notes) 2020* (2020).
- [4] S. Lima-López, E. Farré-Maduell, V. Briva-Iglesias, L. Gasco-Sanchez, M. Krallinger, MEDDOPLACE Shared Task overview: recognition, normalization and classification of locations and patient movement in clinical texts, *Procesamiento del Lenguaje Natural 71* (2023).
- [5] B. Altuna, R. Agerri, S.-E. Lidia, J. J. Saiz, A. Lavelli, B. Magnini, M. Speranza, R. Zanolli, G. Karunakaran, Overview of TESTLINK at IberLEF 2023: Linking Results to Clinical Laboratory Tests and Measurements, *Procesamiento del Lenguaje Natural 71* (2023).
- [6] B. Magnini, B. Altuna, A. Lavelli, A.-L. Minard, M. Speranza, R. Zanolli, European clinical case corpus, in: *European Language Grid: A Language Technology Platform for Multilingual Europe*, Springer International Publishing Cham, 2022, pp. 283–288.
- [7] L. Chiruzzo, M. Agüero-Torales, G. Giménez-Lugo, A. Alvarez, Y. Rodríguez, S. Góngora, T. Solorio, Overview of GUA-SPA at IberLEF 2023: Guarani-Spanish Code Switching Analysis, *Procesamiento del Lenguaje Natural 71* (2023).
- [8] A. M. Mármol-Romero, A. Moreno-Muñoz, F. M. Plaza-del Arco, M.-G. M. Dolores, M. T.

- Martín-Valdivia, L. A. Ureña-López, A. Montejo-Ráez, Overview of MentalRiskES at IberLEF 2023: Early Detection of Mental Disorders Risk in Spanish, *Procesamiento del Lenguaje Natural* 71 (2023).
- [9] H. Jarquín-Vásquez, D. I. Hernández-Farías, L. J. Arellano, H. J. Escalate, L. Villaseñor-Pineda, M. Montes-y Gómez, F. Sanchez-Vega, Overview of DA-VINCIS at IberLEF 2023: Detection of Aggressive and Violent Incidents from Social Media in Spanish, *Procesamiento del Lenguaje Natural* 71 (2023).
- [10] L. J. Arellano, H. J. Escalante, L. Villaseñor-Pineda, M. Montes-y Gómez, F. Sanchez-Vega, Overview of DA-VINCIS at IberLEF 2022: Detection of Aggressive and Violent Incidents from Social Media in Spanish 69 (2022).
- [11] G. Bel-Enguix, H. Gómez-Adorno, G. Sierra, J. Vásquez, S. T. Andersen, S. Ojeda-Trueba, Overview of HOMO-MEX at Iberlef 2023: Hate speech detection in Online Messages directed towards the MEXican Spanish speaking LGBTQ+ population, *Procesamiento del Lenguaje Natural* 71 (2023).
- [12] S. M. Jiménez-Zafra, M. García-Cumbreras, D. García-Baena, J. A. García-Díaz, B. R. Chakravarthi, R. Valencia-García, L. A. Ureña-López, Overview of HOPE at IberLEF 2023: Multilingual Hope Speech Detection, *Procesamiento del Lenguaje Natural* 71 (2023).
- [13] R. Labadie Tamayo, B. Chulvi, P. Rosso, Everybody Hurts, Sometimes Overview of HUrTful HUmour at IberLEF 2023: Detection of Humour Spreading Prejudice in Twitter, *Procesamiento del Lenguaje Natural* 71 (2023).
- [14] P. Moral, G. Marco, J. Gonzalo, J. Carrillo-de Albornoz, I. Gonzalo-Verdugo, Overview of DIPROMATS 2023: automatic detection and characterization of propaganda techniques in messages from diplomats and authorities of world powers, *Procesamiento del Lenguaje Natural* 71 (2023).
- [15] J. A. García-Díaz, S. M. Jiménez-Zafra, M.-T. Martín-Valdivia, F. García-Sánchez, L. A. Ureña-López, R. Valencia-García, Overview of PoliticES at IberLEF 2023: Political Ideology Detection in Spanish Texts, *Procesamiento del Lenguaje Natural* 71 (2023).
- [16] J. A. García-Díaz, S. M. Jiménez-Zafra, M.-T. Martín-Valdivia, F. García-Sánchez, L. A. Ureña-López, R. Valencia-García, Overview of PoliticEs 2022: Spanish Author Profiling for Political Ideology, *Procesamiento del Lenguaje Natural* 69 (2022).
- [17] J. A. García-Díaz, Almela, F. García-Sánchez, G. Alcaraz-Mármol, M. J. Marín, R. Valencia-García, Overview of FinancES 2023: Financial Targeted Sentiment Analysis in Spanish, *Procesamiento del Lenguaje Natural* 71 (2023).
- [18] A. Álvarez Carmona, Miguel A. and Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, V. Muñoz-Sánchez, A. P. López-Monroy, F. Sánchez-Vega, L. Bustio-Martínez, Overview of Rest-Mex at IberLEF 2023: Research on Sentiment Analysis Task for Mexican Tourist Texts, *Procesamiento del Lenguaje Natural* 71 (2023).
- [19] A. Álvarez Carmona, Miguel A. and Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, D. Fajardo-Delgado, R. Guerrero-Rodríguez, L. Bustio-Martínez, Overview of Rest-Mex at IberLEF 2022: Recommendation System, Sentiment Analysis and Covid Semaphore Prediction for Mexican Tourist Texts, *Procesamiento del Lenguaje Natural* 69 (2022).