Brief Overview of TalentCLEF 2025

Luis Gasco[1,*], Hermenegildo Fabregat[1,2], Laura García-Sardiña[1], Paula Estrella[1], Daniel Deniz[1], Alvaro Rodrigo[2] and Rabih Zbib[1]

[1]Avature Machine Learning, Spain
[2]NLP & IR Group at UNED, Madrid, Spain

Abstract

This paper presents a condensed overview of TalentCLEF 2025, the first community evaluation initiative focused on job and skill intelligence in multilingual settings. The campaign attracted 15 participating teams and 280 system submissions from academia and industry across four continents. Analysis of methodological trends reveals a strong reliance on retrieval-based approaches, with selective integration of prompting, re-ranking, and external knowledge. This report highlights key participation insights and methodological patterns that may inform the design of future community challenges in natural language processing for labor market intelligence. TalentCLEF 2025 corpus: https://doi.org/10.5281/zenodo.14002665

Keywords

Natural Language Processing, Human Capital Management, Human Resources, Multilinguality, Cross-linguality, Skill Predictions, Job Title Ranking

1. Introduction

This brief lab report complements the main TalentCLEF 2025 lab overview [1] by providing a focused analysis of participation trends and methodological choices. Rather than revisiting motivation, corpus creation, or evaluation setup in detail, our goal is to identify overarching patterns and extract insights that may inform future shared tasks in this space.

As language technologies have become a strategic component of Human Capital Management (HCM), they are increasingly used to develop systems that semantically analyze resumes and job descriptions [2, 3, 4]. Despite the growing relevance of NLP in this field, progress has been partially limited by the absence of shared benchmarks and standardized evaluation procedures. Open and comparable evaluation frameworks are critical to advance the state of the art, especially in light of persistent challenges such as multilingualism, domain adaptation, and algorithmic bias [5].

TalentCLEF aims to address this gap. As the first community evaluation campaign focused on job and skill intelligence in multilingual contexts, it introduces two tasks that capture the complexity of real-world labor data while emphasizing fairness and cross-lingual applicability. The sections that follow provide a brief overview of task definitions, analyze participation trends, and examine methodological strategies between teams, highlighting key lessons that may guide future shared evaluation initiatives in this area.

2. Tasks

Task A – Multilingual Job Title Matching Participants were asked to retrieve and rank relevant job titles similar to a given job title, in English, Spanish, German, and Chinese. The evaluation covered monolingual and cross-lingual scenarios, as well as gender-based evaluation. The official evaluation metrics are Mean Average Precision (MAP) and Rank Biased Overlap (RBO).

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain
*Corresponding author.
machinelearning@avature.net (L. Gasco)
0000-0002-4976-9879 (L. Gasco); 0000-0001-9820-2150 (H. Fabregat); 0000-0003-4592-8884 (L. García-Sardiña); 0000-0002-0313-2127 (D. Deniz); 0000-0002-6331-4117 (A. Rodrigo); 0000-0002-7140-3048 (R. Zbib)


Task B – Job Title-Based Skill Prediction Given a job title, the systems had to detect relevant skills using a subset of ESCO skills. In this task only English was considered and the official evaluation metric was the Mean Average Precision.

Both corpora were built from real job offers and applications, manually annotated, and are publicly available in Zenodo[1]. The official benchmark results can be accessed via the task website[2], and the evaluation platform remains available on Codabench for Task A [3] and Task B [4].

3. Participation

TalentCLEF 2025 included 15 teams that developed a total of 280 system runs for the task. In particular, Task A had 12 participants and Task B had 8 teams.

Figure 1 shows a global map with the location of participating institutions, as well as their participation in Task A and/or B. Teams came from Europe, America, Africa, and Asia. Most of them came from universities and research centers, which shows that there is a strong interest from academia in working with NLP applied to Human Resources. However, with the exception of TechWolf, industry participation was limited, indicating that companies may still be reluctant to engage openly in this kind of shared evaluation initiative.

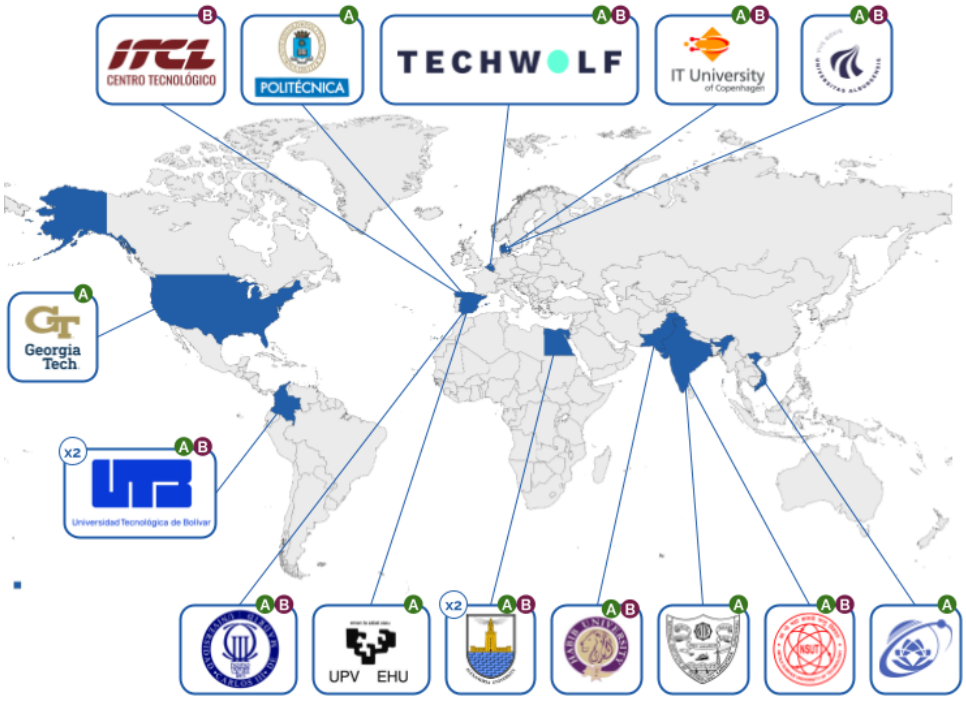


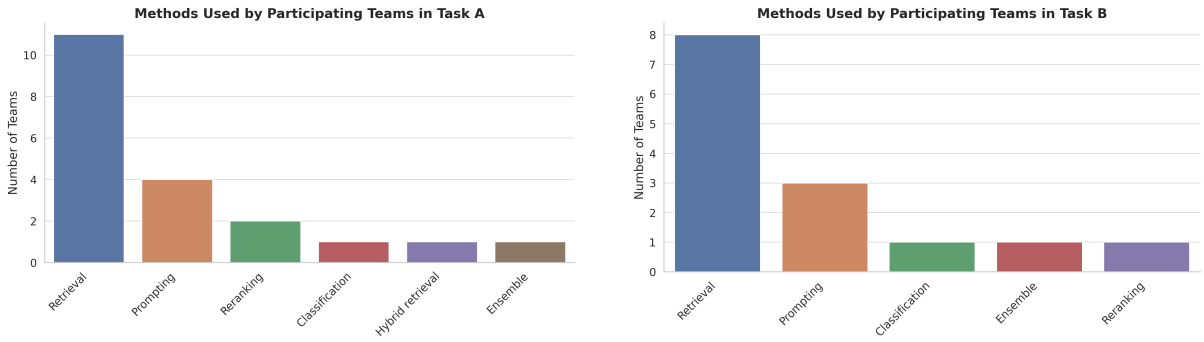
Figure 1: Geographical distribution of the institutions participating in TalentCLEF 2025. The figure also shows whether each institution submitted systems to Task A, Task B, or both. In some cases, more than one team from the same institution participated.

4. Methodologies

An analysis of the technologies used in the submitted systems highlights the central role of retrieval-based methods in both Task A and Task B for identifying relevant elements given a query, despite the order not being relevant in the evaluation. As shown in Figure 2, almost all participating teams relied on

[1]TalentCLEF Zenodo: https://doi.org/10.5281/zenodo.14002665
[2]Full TalentCLEF 2025 results: https://talentclef.github.io/talentclef/docs/talentclef-2025/results
[3]Task A Codabench: https://www.codabench.org/competitions/5842/
[4]Task B Codabench: https://www.codabench.org/competitions/7059/

retrieval mechanisms, while only a subset incorporated complementary techniques such as prompting with large language models (LLM) or re-ranking strategies. This trend underscores the foundational role of retrieval in these relevance tasks as used in previously published literature [6, 7]. Although the tasks did not explicitly provide contextual information, many participants opted to enrich their systems using external knowledge sources, such as controlled vocabularies like ESCO. Moreover, some teams used large language models to generate synthetic data, in order to enhance the retrieval of relevant elements and improve overall performance.



(a) Methods used in Task A

(b) Methods used in Task B

Figure 2: Overview of methods used by participating systems across Task A and Task B.

5. Conclusions

The first edition of TalentCLEF can be considered a success in terms of participation, attracting interest from both academia, driven by the development of NLP technologies in underexplored areas, and industry, which showed a willingness to openly share how their systems perform in open benchmark settings. This level of involvement provides a strong foundation for organizing future editions of TalentCLEF, with opportunities to further advance complex tasks such as Task B and to introduce new challenges aimed at evaluating higher-risk systems, including those based on generative AI techniques.

Acknowledgments

The authors express their sincere gratitude to the members of the scientific committee who contributed to the dissemination and visibility of the TalentCLEF initiative. Their support was instrumental in reaching a diverse audience in academia and industry and in promoting the development of robust and impactful NLP technologies for the labor market. The scientific committee members of this task are:

- Eneko Agirre – Full Professor at the University of the Basque Country (UPV/EHU), ACL Fellow
- David Camacho – Full Professor at the Technical University of Madrid (UPM)
- Maria R. Costa-Jussà – Research Scientist at Meta AI
- Debora Nozza – Assistant Professor at Bocconi University
- Jens-Joris Decorte – Lead AI Scientist at TechWolf
- David Graus – Lead Data Scientist at Randstad Group
- Mesutt Kayaa – Postdoctoral Researcher at Jobindex A/S and IT University of Copenhagen
- Jan Luts – Senior Data Scientist at NTT Data and ESCO
- Elena Montiel-Ponsoda – Professor at the Technical University of Madrid (UPM), AI4Labour project
- Javier Huertas Tato – Assistant Professor at the Technical University of Madrid (UPM)
- Patricia Martín Chozas – Postdoctoral Researcher at the Ontology Engineering Group (UPM), AI4Labour project

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4o to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] L. Gasco, H. Fabregat, L. García-Sardiña, P. Estrella, D. Deniz, A. Rodrigo, R. Zbib, Overview of the TalentCLEF 2025: Skill and Job Title Intelligence for Human Capital Management, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2025.

[2] F. Retyk, H. Fabregat, J. Aizpuru, M. Taglio, R. Zbib, Résumé parsing as hierarchical sequence labeling: An empirical study, in: M. Kaya, T. Bogers, D. Graus, C. Johnson, J. Decorte (Eds.), Proceedings of the 3rd Workshop on Recommender Systems for Human Resources (RecSys in HR 2023) co-located with the 17th ACM Conference on Recommender Systems (RecSys 2023), Singapore, Singapore, 18th-22nd September 2023, volume 3490 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023. URL: https://ceur-ws.org/Vol-3490/RecSysHR2023-paper_10.pdf.

[3] F. Retyk, L. Gascó, C. P. Carrino, D. Deniz, R. Zbib, MELO: an evaluation benchmark for multilingual entity linking of occupations, in: M. Kaya, T. Bogers, D. Graus, C. Johnson, J. Decorte, T. D. Bie (Eds.), Proceedings of the 4th Workshop on Recommender Systems for Human Resources (RecSys-in-HR 2024) co-located with the 18th ACM Conference on Recommender Systems (RecSys 2024), Bari, Italy, 14th-18th October 2024, volume 3788 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024. URL: https://ceur-ws.org/Vol-3788/RecSysHR2024-paper_2.pdf.

[4] H. Fabregat, R. Poves, L. L. Alvarez, F. Retyk, L. García-Sardiña, R. Zbib, Inductive graph neural network for job-skill framework analysis, Procesamiento del Lenguaje Natural 73 (2024) 83–94. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6602.

[5] L. Gasco, H. Fabregat, L. García-Sardiña, D. Deniz, A. Rodrigo, P. Estrella, R. Zbib, TalentCLEF at CLEF2025: Skill and Job Title Intelligence for Human Capital Management, in: European Conference on Information Retrieval, Springer, 2025, pp. 479–486.

[6] D. Deniz, F. Retyk, L. García-Sardiña, H. Fabregat, L. Gascó, R. Zbib, Combined unsupervised and contrastive learning for multilingual job recommendation, in: M. Kaya, T. Bogers, D. Graus, C. Johnson, J. Decorte, T. D. Bie (Eds.), Proceedings of the 4th Workshop on Recommender Systems for Human Resources (RecSys-in-HR 2024) co-located with the 18th ACM Conference on Recommender Systems (RecSys 2024), Bari, Italy, 14th-18th October 2024, volume 3788 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024. URL: https://ceur-ws.org/Vol-3788/RecSysHR2024-paper_3.pdf.

[7] R. Zbib, L. L. Alvarez, F. Retyk, R. Poves, J. Aizpuru, H. Fabregat, V. Simkus, E. G. Casademont, Learning job titles similarity from noisy skill labels, CoRR abs/2207.00494 (2022). URL: https://doi.org/10.48550/arXiv.2207.00494. doi:10.48550/ARXIV.2207.00494. arXiv:2207.00494.