UDII-UPM at TalentCLEF 2025: Task A-Multilingual Job Title Matching

Javier Rodríguez-Vidal[1,*,†], Ascensión López-Vargas[1,†], Pablo Manuel Vigara Gallego[1,†], Francisco Javier Del Álamo[1,†] and Ángel García-Beltrán[1,†]

[1]*Departamento de Automática, Ingeniería Eléctrica y Electrónica e Informática Industrial, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, 28006 Madrid, Spain*

Abstract

This paper describes the participation of the Unidad Docente de Informática Industrial at the Universidad Politécnica de Madrid (UDII-UPM) in TalentCLEF 2025, a competitive evaluation campaign focused on Human Resources. We addressed Task A: Multilingual Job Title Matching, which aims to develop systems capable of identifying and ranking job titles that are most similar to a given query title. Our approach is based on the fusion of semantic representations generated by two different multilingual embedding models. First, we compute pairwise similarities between query and candidate job titles. Then, we apply a weighted combination of the resulting similarity scores to produce a final ranked list of job titles for each query. The proposed method is simple, language-adaptable and does not require fine-tuning. Official results confirm that our system is competitive although there is still room for improvement in the ranking quality and model alignment across languages.

Keywords

Human Capital Management, Human Resources, Embeddings, Ranking, Information Retrieval

1. Introduction

The emergence of new job roles and the evolution of existing ones (driven by rising levels of automation, Artificial Intelligence, and the widespread adoption of remote work) have made it increasingly necessary to define highly specialized professional profiles. However, recruiting for these roles has become progressively more challenging.

Simultaneously, the rapid progress of language-based technologies is transforming Human Capital Management and Human Resources, enabling the identification of the most suitable candidates for specific positions from vast volumes of data.

This paper presents the participation of the Unidad Docente de Informática Industrial at the Universidad Politécnica de Madrid (UDII-UPM) in TalentCLEF 2025 Task A: Multilingual Job Title Matching [1]. The main objective of this task[1] is to "develop systems that can identify and rank job titles most similar to a given job title. For each job title in a provided test set, participants must generate a ranked list of similar job titles from a specified knowledge base." This is a multilingual challenge, as the systems must be able to adapt to English, Spanish, German, and optionally, Chinese.

This work has two main goals: i) to assess whether pre-trained systems perform well on tasks in different languages without the need for fine-tuning, and ii) to establish a baseline for future developments.

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain

*Corresponding author.

†These authors contributed equally.

javier.rodriguez.vidal@upm.es (J. Rodríguez-Vidal); a.lvargas@upm.es (A. López-Vargas); pm.vigara@upm.es (P. M. V. Gallego); fjalamo@etsii.upm.es (F. J. D. Álamo); agarcia@etsii.upm.es (Á. García-Beltrán)

0000-0002-9006-9639 (J. Rodríguez-Vidal); 0000-0002-8492-5929 (A. López-Vargas); 0009-0003-6572-6697 (P. M. V. Gallego); 0000-0003-0595-2134 (F. J. D. Álamo); 0000-0003-1900-0222 (Á. García-Beltrán)

[1]https://talentclef.github.io/talentclef/docs/talentclef-2025/task-summary/

2. Experimental Framework

In this section, we give an overview of the Job Title Matching dataset, the methods used, how the run delivered was set up, and the evaluation metric used in the lab.

2.1. Multilingual Job Title Matching corpus

The corpus [2] consisted of training, validation and test job title data divided in three main languages: English, Spanish and German; validation and test also contains information in Chinese. The training data contains 4 columns: i) the ISCO id representing the group to which the job identifier belongs; ii) the ESCO id indicating the origin of the pair's job titles; iii) the first job title in the pair and iv) a second job title related to the previous one. The statistics of this part of the dataset are shown in Table 1.

Table 1
Description of training data

Language	Number of Data
English	28880
Spanish	20724
German	23023

The validation data are divided into four languages, the three in the training data and Chinese. This set contains: i) a file that contains the job title used as a query and the associated id (*queries*); ii) the job title in the corpus and its id (*corpus_elements*) and iii) the relationships between the query and the corpus elements (*qrels*). The statistics of the validation part of the dataset are shown in Table 2.

Table 2
Description of validation data

Language	Queries	Corpus_elements	Qrels
English	105	2619	2420
Spanish	185	4661	7579
German	203	4729	8417
Chinese	103	2513	2319

The test data has four languages, as the validation set, and the provided files: i) contains the job title used as a query and the associated id (*queries*) and ii) contains the job title in the corpus and its id (*corpus_elements*). The statistics of the test part of the dataset are shown in Table 3.

Table 3
Description of test data

Language	Queries	Corpus_elements
English	117	770
Spanish	192	1232
German	227	1510
Chinese	117	770

2.2. Methods

This study uses two learning algorithms as a starting point for future work. In particular, the following algorithms have been used:

- **Masked and Permuted Language Modeling (MPNET)** [3] based on the BERT architecture [4] and uses Siamese and Triplet network structures to extract sentence embeddings that can be compared using cosine similarity. The Siamese network is used to determine whether two sentences are similar, while the Triplet network uses three sentences: an anchor, a positive (similar), and a negative (dissimilar). The model compares the distance between the anchor and the positive sentence and between the anchor and the negative one. It learns to bring similar sentences closer together and push dissimilar ones farther apart.
- **DistilUSE** [3]: is a multilingual sentence embedding model based on the DistilBERT [5] architecture, a lighter and faster version of BERT. It supports over 50 languages and is trained to produce semantically meaningful sentence embeddings that can be compared using cosine similarity. While it does not rely on Siamese or Triplet networks during training, it performs well in semantic similarity tasks and multilingual sentence matching, especially when computational efficiency is important.

These models were used via HuggingFace[2]

2.3. Setting up the run

Our proposal consists of the weighted combination of the similarities produced by the models mentioned in section 2.2. To achieve this, each model individually generates embeddings for the provided texts. Then, a similarity matrix is computed between the query texts and the corpus elements, and these similarities are fused according to the following equation:

$$similarity\_fusion = \alpha \times similarity\_mpnet + (1 - \alpha) \times similarity\_distiluse \quad (1)$$

where, α are the fusion weights, empirically determined for each language, described in Table 4; *similarity_mpnet* is the similarity between queries and corpus texts using mpnet embeddings and *similarity_distiluse* is the similarity between queries and corpus texts using distiluse embeddings.

Table 4
Weights used according to each language

Language	Weights
English	0.80
Spanish	0.75
German	0.80
Chinese	0.60

Finally, the output ranking of documents is generated by sorting them in descending order of similarity.

2.4. Evaluation Metric

To evaluate our approach during the development phase, we used the official evaluation script provided by the organizers.[3] The final performance was assessed using the *Mean Average Precision* (MAP) metric [6].

3. Results & Discussion

Table 5 summarizes the official results, based on the test set of the Task, providing only the results of the system ranked in the first place, the system with the best cross-lingual predictions (in average), our results and the baseline provided by the organizers.

[2] https://huggingface.co/
[3] https://github.com/TalentCLEF/talentclef25_evaluation_script

Table 5
Official results

Language	en-en	es-es	de-de	zh-zh	en-es	en-de	en-zh	Overall
Best system (ranabarakat)	0.559	0.527	0.516	0.508	n/a	n/a	n/a	0.53
Best C-L (pjmathematician)	0.563	0.507	0.476	0.516	0.525	0.504	0.524	0.52
udii-upm	0.448	0.415	0.377	0.451	0.383	0.375	0.426	0.41
Baseline (TalentCLEF)	0.408	0.348	0.324	0.380	0.335	0.345	n/a	0.36

Based on the results, our system consistently obtains better results than the baseline; however, these results are far from the system that achieved the best results. This may happen because our system does not perform a training step prior to the generation of the embeddings so the methods used here are not adapted to the task. Also, as can be seen from the results, the language with which we have encountered the most problems is German, which may be due to the fact that it is the language with the most number of queries and corpus texts. If we observed the rest of the official results provided in Table 5 the other systems also had problems with this language.

4. Future Work

In this work udii-upm team provides information of its participation at TalentCLEF 2025, Task A-Multilingual Job Title Matching. The system presented in this task provides a good starting point for future developments which includes: i) train the models with the training data provided in the corpus; ii) explore in depth the problems presented in the German language in order to improve the results and iii) to explore of monolingual models instead of multilingual ones.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4o mini in order to: Grammar and spelling check.

References

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

[2] L. Gascó, F. M. Hermenegildo, G.-S. Laura, D. C. Daniel, P. Estrella, R. Alvaro, Z. Rabih, Talentclef 2025 corpus: Skill and job title intelligence for human capital management, 2025. URL: https://doi.org/10.5281/zenodo.15292308. doi:10.5281/zenodo.15292308.

[3] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019. URL: http://arxiv.org/abs/1908.10084.

[4] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423/. doi:10.18653/v1/N19-1423.

[5] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, ArXiv abs/1910.01108 (2019).

[6] H. Schütze, C. D. Manning, P. Raghavan, Introduction to information retrieval, volume 39, Cambridge University Press Cambridge, 2008.