VerbaNexAI at TalentCLEF 2026: A Hybrid Multilingual Retrieval and Approach for Contextualized Job-Person Matching

Working notes paper for the TalentCLEF Lab at CLEF 2026

Melissa Moreno Novoa, Juan Carlos Martinez-Santos and Edwin Puertas

Universidad Tecnologica de Bolivar, Cartagena, Colombia

Abstract

This paper presents the participation of the VerbaNex AI team in TalentCLEF 2026 Task A: Contextualized Job-Person Matching, a shared task focused on retrieving and ranking the most relevant candidates for job vacancies in multilingual environments. To address this challenge, we propose a hybrid retrieval architecture that combines dense semantic retrieval based on multilingual embeddings with lexical retrieval using BM25, followed by a neural re-ranking stage employing a multilingual cross-encoder.

The proposed framework includes document preprocessing, job posting enrichment, hybrid candidate retrieval, and contextualized reranking. Candidate profiles and job descriptions are represented using the BGE-M3 multilingual bi-encoder, enabling semantic matching across English and Spanish. Retrieval scores obtained from dense and lexical approaches are normalized and combined through weighted fusion, while a cross-encoder refines the final ranking by modeling deeper interactions between job requirements and candidate qualifications.

Experiments were conducted using the official TalentCLEF 2026 Task A dataset under monolingual and cross-lingual retrieval settings. The proposed system achieved an average MAP of 0.560, obtaining MAP scores of 0.558 for English-English retrieval and 0.554 for Spanish-Spanish retrieval. Additionally, the system reached MRR values of 0.859 and 0.840 in the respective monolingual scenarios. Results demonstrate the effectiveness of combining semantic and lexical retrieval strategies for contextualized talent matching, while also highlighting the remaining challenges associated with cross-lingual candidate retrieval and ranking.

Keywords

Job-Person Matching, Information Retrieval, Multilingual NLP, Candidate Ranking

1. Introduction

The labor market has undergone a significant digital transformation in recent years, driven by automated recruitment systems, remote hiring processes, and the growing demand for specialized talent. In this context, Artificial Intelligence (AI) and Natural Language Processing (NLP) have become essential technologies for efficiently analyzing large volumes of resumes and job descriptions [1] [2].

These technologies enable the generation of semantic representations of resumes and job descriptions, facilitating the comparison of skills, experience, and professional requirements beyond keyword-based matching. Furthermore, modern systems integrate retrieval and ranking mechanisms to identify the most relevant profiles for a given job vacancy [3][4].

Despite the advances achieved by Transformer-based models, significant challenges remain in the accurate identification of candidate profiles due to skill overlap among similar occupations. Moreover, the limited availability of labeled data and privacy constraints hinder the development of more robust systems. Likewise, adaptation to different industries and the preservation of semantic context in multilingual scenarios continue to be open research problems [5][6].

To address these challenges, TalentCLEF has emerged as an initiative for evaluating intelligent systems aimed at talent management. In particular, Task A: Contextualized Job-Person Matching focuses on identifying and ranking relevant candidates for job opportunities in multilingual English

CLEF 2026 Working Notes, 21 – 24 September 2026, Jena, Germany
memoreno@utb.edu.co (M. M. Novoa); jcmartinezs@utb.edu.co (J. C. Martinez-Santos); epuerta@utb.edu.co (E. Puertas)
0000-0002-1748-2208 (M. M. Novoa); 0000-0003-2755-0718 (J. C. Martinez-Santos); 0000-0002-0758-1851 (E. Puertas)

and Spanish environments. System performance is evaluated using Information Retrieval metrics, with Mean Average Precision (MAP) serving as the primary evaluation metric, complemented by MRR and Precision@K [7][8].

Within this context, this paper presents the participation of the VerbaNex AI team in TalentCLEF Task A through a hybrid retrieval system that combines multilingual dense retrieval using semantic embeddings, lexical retrieval based on BM25, and cross-encoder reranking to improve the identification of relevant candidates. The objective was to enhance contextualized job-candidate matching while evaluating system performance using standard information retrieval metrics.

2. Related Work

Traditional Applicant Tracking Systems (ATS) primarily relied on exact term matching, which primarily relied on search strategies based on exact term matching, supported by skill dictionaries, Boolean filters, and heuristic rules to classify candidates [9]. Although these systems facilitated the initial automation of recruitment processes, their strong dependence on lexical matching and the need for human intervention to interpret complex profiles limited their ability to capture semantic relationships and adapt to evolving labor market demands [6]. These limitations motivated the development of more advanced approaches capable of understanding the contextual meaning of employment-related information beyond simple textual matching.

The incorporation of NLP techniques represented a significant advancement by enabling the automatic analysis of resume and job posting content. Models such as Word2Vec, Doc2Vec, and later Sentence-BERT made it possible to project candidates and job vacancies into dense vector spaces where similarity could be computed using metrics such as cosine similarity. This facilitated the identification of equivalent competencies described using different vocabularies. Nevertheless, many of these solutions continued to rely on static representations or models trained on general-domain corpora, limiting their ability to capture complex contextual relationships and specialized terminology, particularly in multilingual or highly specialized scenarios [10][11].

More recently, Transformer models and Large Language Models (LLMs) have driven a new generation of job-matching systems by providing deep contextual representations of resumes and job descriptions [12, 13]. These models enable the capture of complex relationships among experience, skills, and job requirements, thereby improving the semantic understanding of the matching process. However, their adoption introduces challenges related to high computational costs, scalability in environments with large volumes of data, and the ability to generalize effectively across different domains and labor markets [14].

In parallel, Information Retrieval (IR) has assumed a central role in modern job-matching systems, where architectures based on bi-encoders and dual encoders enable the implementation of dense retrieval strategies for efficiently retrieving and ranking relevant candidates [15, 16]. Although these approaches have improved retrieval performance, challenges related to multilinguality, the representation of implicit skills, and accurate candidate ranking remain unresolved, motivating evaluation initiatives such as TalentCLEF.

3. Experimental Framework

This section presents the experimental framework proposed to address the contextualized job-person matching task. The approach integrates text preprocessing, query enrichment, hybrid retrieval, and neural reranking techniques to identify and rank the most relevant candidates for each job vacancy. The system was designed to operate in both monolingual and multilingual scenarios, enabling the evaluation of its retrieval capabilities in English and Spanish.

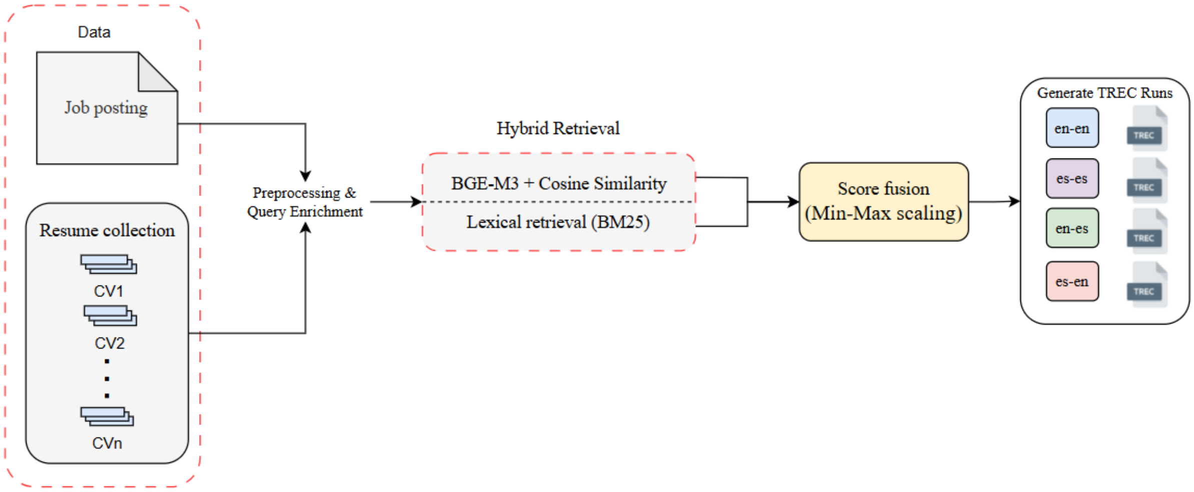

Figure 1: Overview of the proposed hybrid retrieval pipeline.

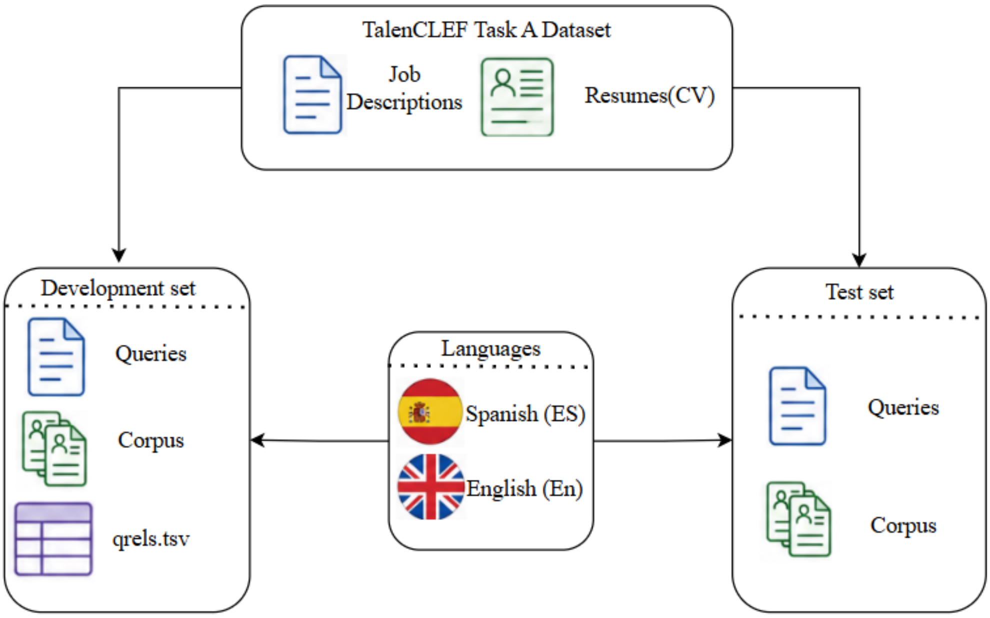

Figure 2: Structure of the TalentCLEF Task A dataset.

3.1. Method Overview

Figure 1 illustrates the overall architecture proposed for the contextualized job-person matching task. The approach follows a two-stage hybrid retrieval strategy. Initially, job postings and resumes undergo a preprocessing and text enrichment process to improve the semantic representation of information. A hybrid retrieval mechanism is then applied, combining dense representations generated by a multilingual bi-encoder with lexical retrieval based on BM25. The retrieved candidates are subsequently reranked using a cross-encoder, producing a final ranking of the most relevant profiles for each vacancy.

3.2. Dataset

The experiments were conducted using the dataset provided by TalentCLEF Task A, which focuses on the Contextualized Job-Person Matching problem. The corpus consists of job postings and resumes organized into separate collections for English and Spanish. Additionally, the development set includes relevance judgments that enable the evaluation of ranking quality during the validation phase.

To evaluate the generalization capability of the system, three retrieval scenarios were considered: English-English (EN-EN), Spanish-Spanish (ES-ES), and cross-lingual English-Spanish retrieval (EN-ES). This configuration allows the analysis of both monolingual performance and the model's ability to perform matching across different languages Figure 2 .

3.3. Data Preprocessing

Before the retrieval stage, all documents underwent a text normalization process. Elements considered informational noise, including email addresses, web links, and phone numbers, were removed. The text was converted to lowercase, and special characters deemed irrelevant to the retrieval task were eliminated.

Unlike more aggressive cleaning procedures, technical expressions frequently used in professional contexts, such as programming languages, frameworks, and specialized technologies, were preserved. This decision aims to retain information relevant to the identification of specific job-related competencies and to avoid the loss of important semantic signals during the matching process.

To enrich the representation of job postings, a structured query expansion strategy was applied. Instead of using an external generative model, each job posting was reformulated into an enriched textual query by explicitly adding sections related to core skill requirements, experience preferences, educational background requirements, and implicit candidate preferences. These sections incorporated task relevant information, including technical skills, tools, technologies, programming languages, certifications, years of experience, previous responsibilities, academic background, transferable skills, domain expertise, and job–resume alignment.

This enrichment strategy aimed to make implicit job requirements more explicit for both dense and lexical retrieval. The enriched query was used as input for the BGE-M3 dense retriever and was also tokenized for BM25 retrieval, while candidate resumes were cleaned and indexed without additional expansion.

3.4. Retrieval Architecture

The retrieval stage was designed under a hybrid approach that combines dense semantic retrieval and traditional lexical retrieval. For the semantic representation of job postings and resumes, the multilingual BGE-M3 model was employed, a bi-encoder capable of generating dense embeddings within a shared vector space across multiple languages. This characteristic is particularly relevant for multilingual job-matching scenarios, where queries and documents may be written in different languages.

The vectors generated by the bi-encoder were compared using cosine similarity to estimate the degree of semantic correspondence between each job posting and candidate profile. A BM25 index was constructed using the preprocessed resumes, allowing the capture of exact lexical matches between terms appearing in job postings and corpus documents. This combination aims to leverage the complementary strengths of both approaches: the semantic capabilities of dense embeddings and the lexical precision provided by BM25.

Since both retrieval strategies generate scores on different scales, the results were normalized using Min-Max Scaling. Subsequently, the scores were combined through a weighted linear strategy defined as:

$$Score_{Hybrid} = \alpha \cdot Score_{Dense} + (1 - \alpha) \cdot Score_{BM25} \tag{1}$$

where α controls the relative contribution of dense semantic retrieval and lexical BM25 retrieval. In our experiments, $\alpha = 0.60$ was selected based on preliminary validation on the development set. The resulting hybrid score was used to rank candidates before the reranking stage, retaining the top 700 candidates for subsequent processing.

3.5. Re-ranking Stage

Although hybrid retrieval enables the identification of potentially relevant candidates, the independent representations generated by bi-encoders limit the ability to model deep interactions between queries and documents. To address this limitation, a second reranking stage based on a multilingual cross-encoder was incorporated.

Unlike bi-encoders, the cross-encoder simultaneously processes the job posting and resume within a single input sequence, enabling the modeling of more complex contextual relationships between both texts. In this work, the Cross-Encoder mMiniLMv2 model was employed, which is widely used in retrieval and reranking tasks due to its balance between effectiveness and computational efficiency.

During this stage, each retrieved candidate was re-evaluated by the cross-encoder, obtaining a new relevance score. The scores were normalized and combined with the initial score generated by the hybrid system using the following expression:

$$Score_{Final} = \lambda \cdot Score_{Cross} + (1 - \lambda) \cdot Score_{Hybrid} \quad (2)$$

where λ controls the influence of the reranking model. In the experiments, $\lambda = 0.95$ was used, assigning greater weight to the contextual evaluation produced by the cross-encoder. Finally, the top 100 highest-scoring candidates were retained to generate the final ranking for each query.

3.6. Experimental Settings

All experiments were implemented using Python together with the Sentence Transformers, PyTorch, Rank-BM25, and Ranx libraries. Processing was executed on an NVIDIA GeForce RTX 4070 Laptop GPU, accelerating both embedding generation and neural reranking.

The BGE-M3 model was employed as the primary dense retriever, while BM25 served as the complementary lexical retrieval mechanism. For the reranking stage, the Cross-Encoder mMiniLMv2 model was utilized. Embedding generation and cross-encoder predictions were performed using a batch size of 8 samples. Furthermore, the system initially retrieved the top 700 ranked documents per query and subsequently retained only the top 100 candidates after the reranking stage.

Evaluation was conducted on the development set using the relevance judgments provided by the task organizers. Following the official TalentCLEF protocol, the metrics MAP, MRR, nDCG@10, Precision@5, and Precision@10 were employed. These metrics assess both the overall ranking quality and the system's ability to position relevant candidates at the top of the retrieved results.

Finally, the generated rankings were exported in TREC format, automatically validating the required structure before generating the final files used for official evaluation. This strategy ensures experiment reproducibility and compatibility with the TalentCLEF evaluation infrastructure.

4. Results and Discussion

Table 1 presents the official results obtained by the proposed system in TalentCLEF 2026 Task A. The proposed system achieved an average MAP of 0.56 across the evaluated language settings, obtaining MAP scores of 0.558 and 0.554 for the en-en and es-es scenarios, respectively. The highest effectiveness was observed in monolingual retrieval, where both language configurations produced similar results, indicating a consistent ability to identify relevant candidates within the same language.

Regarding ranking quality, the system achieved MRR values of 0.859 for en-en and 0.840 for es-es, suggesting that relevant candidates were frequently retrieved within the first positions of the ranking. In contrast, the cross-lingual scenario (en-es) proved considerably more challenging, obtaining a MAP of 0.358 and an MRR of 0.633. This decrease highlights the difficulties associated with matching candidate profiles and job descriptions across languages, where semantic equivalence may not always be explicitly represented.

As shown in Table 2, the proposed system consistently outperformed the official TalentCLEF baseline in the average MAP as well as in the two monolingual retrieval settings. The largest improvements were

Table 1
Official retrieval performance obtained on the TalentCLEF 2026 Task A benchmark.

Language Pair	MAP	MRR
EN–EN	0.558	0.859
ES–ES	0.554	0.840
EN–ES	0.358	0.633
Average (EN, ES)	**0.560**	–

Table 2
Comparison between the official TalentCLEF baseline and the proposed system.

System	Avg. MAP	EN–EN	ES–ES	EN–ES	MRR EN–EN	MRR ES–ES	MRR EN–ES
Baseline	0.380	0.399	0.365	0.355	0.744	0.688	0.712
VerbaNexAI	**0.560**	**0.558**	**0.554**	**0.358**	**0.859**	**0.840**	0.633

observed for EN–EN and ES–ES retrieval, indicating that the combination of dense semantic retrieval, lexical retrieval, and cross-encoder reranking effectively improves candidate ranking within the same language. In the cross-lingual EN–ES scenario, The proposed approach achieved a slightly higher MAP while obtaining a slightly lower MRR than the official baseline in the EN–ES retrieval scenario while obtaining a slightly lower MRR than the baseline, highlighting that multilingual retrieval remains the most challenging setting and an important direction for future improvements.

5. Future Work

Although the obtained results demonstrate the effectiveness of semantic representations for contextualized job-person matching, significant challenges remain in multilingual retrieval scenarios. As future work, we propose exploring more robust cross-lingual semantic alignment strategies through embedding models specifically designed for cross-lingual retrieval and domain adaptation techniques tailored to employment-related contexts. Furthermore, it would be valuable to evaluate more advanced reranking architectures based on LLMs, which are capable of capturing implicit relationships among skills, professional experience, and job requirements.

Another promising research direction involves incorporating structured information derived from occupational taxonomies and knowledge bases, such as ESCO, to enrich skill representations and improve the identification of semantically related competencies. In addition, future evaluations could explore hybrid approaches that combine dense retrieval, query expansion, and contextual reasoning to enhance performance in multilingual and low lexical-overlap scenarios. These improvements would contribute to the development of more accurate, scalable, and adaptable talent search systems for international labor markets.

Acknowledgments

The authors express their gratitude to the Call 933 "Training in National Doctorates with a Territorial, Ethnic and Gender Focus in the Framework of the Mission Policy — 2023" of the Ministry of Science, Technology and Innovation (Minciencia). In addition, we thank the team of the Artificial Intelligence Laboratory VerbaNex https://github.com/VerbaNexAI, affiliated with the UTB, for their contributions to this project.

Declaration on Generative AI

During the preparation of this manuscript, ChatGPT (OpenAI) was used for the revision of translations into English, as well as for grammatical and spelling correction. After using this tool, the content was reviewed and edited as necessary, and full responsibility for the content of the publication is assumed.

References

[1] D. Spoladore, S. Criscuolo, F. Isgrò, Representing jobs and job seekers in ai-based recommender systems: A literature review, Engineering Applications of Artificial Intelligence 174 (2026) 114450. URL: https://www.sciencedirect.com/science/article/pii/S0952197626007311. doi:https://doi.org/10.1016/j.engappai.2026.114450.

[2] R. V. K. Bevara, N. R. Mannuru, S. P. Karedla, B. Lund, T. Xiao, H. Pasem, S. C. Dronavalli, S. Rupeshkumar, Resume2vec: Transforming applicant tracking systems with intelligent resume embeddings for precise candidate matching, Electronics 14 (2025). URL: https://www.mdpi.com/2079-9292/14/4/794. doi:10.3390/electronics14040794.

[3] L. Aminou, A. Daaif, M. Soulami, A. Chalfaouat, M. Youssfi, Are cutting-edge technologies transforming classical approaches in automated hiring systems ? a systematic literature review, Systems and Soft Computing 7 (2025) 200421. URL: https://www.sciencedirect.com/science/article/pii/S2772941925002406. doi:https://doi.org/10.1016/j.sasc.2025.200421.

[4] H. Li, X. Tang, Q. Liu, B. Liu, S. Zhao, Enhancing intelligent recruitment with generative pretrained transformer and hierarchical graph neural networks, Journal of Organizational and End User Computing 37 (2025) 1–24. doi:10.4018/JOEUC.377613.

[5] J. Li, B. Xu, Z. Chen, C. Xu, M. Chen, S. Liu, Y. Zhou, Z. Wen, Enhancing talent search ranking with role-aware expert mixtures and llm-based fine-grained job descriptions, in: Proceedings of the ACM Web Conference, 2025, pp. 185–194.

[6] A. Bondielli, F. Marcelloni, On the use of summarization and transformer architectures for profiling résumés, Expert Systems with Applications 184 (2021). doi:10.1016/j.eswa.2021.115521.

[7] L. Gasco, H. Fabregat, L. García-Sardiña, P. Estrella, W. Veys, C. P. Carrino, M. De Lange, D. Deniz Cerpa, A. Rodrigo, J.-J. Decorte, R. Zbib, Overview of the talentclef 2026: Skill and job title intelligence for human capital management, in: International Conference of the Cross-Language Evaluation Forum for European Languages, 2026.

[8] H. Fabregat, L. García-Sardiña, P. Estrella, L. Gasco, C. P. Carrino, D. Deniz Cerpa, A. Rodrigo, R. Zbib, Overview of talentclef 2026: Task a – contextualized job-person matching, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2026), 2026.

[9] J. Rosenberger, L. Wolfrum, S. Weinzierl, M. Kraus, P. Zschech, Careerbert: Matching resumes to esco jobs in a shared embedding space for generic job recommendations, Expert Systems with Applications 275 (2025). doi:10.1016/j.eswa.2025.127043.

[10] R. Alonso, D. Dessí, A. Meloni, D. R. Recupero, A novel approach for job matching and skill recommendation using transformers and the o*net database, Big Data Research 39 (2025). doi:10.1016/j.bdr.2025.100509.

[11] Z. Wang, W. Wei, C. Xu, J. Xu, X. L. Mao, Person-job fit estimation from candidate profile and related recruitment history with co-attention neural networks, Neurocomputing 501 (2022) 14–24. doi:10.1016/j.neucom.2022.06.012.

[12] M. Vyas, A. Chakraborty, V. Gupta, Jobmatchai: An intelligent job matching platform using knowledge graphs, semantic search and explainable ai, arXiv preprint (2026). URL: http://arxiv.org/abs/2603.14558.

[13] X. Yu, R. Xu, C. Xue, J. Zhang, X. Ma, Z. Yu, Confit v2: Improving resume-job matching using hypothetical resume embedding and runner-up hard-negative mining, arXiv preprint (2025). URL: http://arxiv.org/abs/2502.12361.

[14] E. D. Campion, M. A. Campion, Impact of machine learning on personnel selection, Organizational Dynamics 53 (2024) 101035. doi:10.1016/j.orgdyn.2024.101035.
[15] D. Lavi, Learning to match job candidates using multilingual bi-encoder bert, in: Fifteenth ACM Conference on Recommender Systems, RecSys '21, ACM, 2021, p. 565–566. URL: http://dx.doi.org/10.1145/3460231.3474609. doi:10.1145/3460231.3474609.
[16] 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, 2025, pp. 464–485.

A. Online Resources

The implementation of the proposed system is available at:

- GitHub