LSI_UNED at BioNNE-R 2026: Improving Biomedical Relation Extraction Through Candidate Space Reduction

Alicia Ramirez-Arrabe[1,*], Juan Martinez-Romo[1,2] and Andres Duque[1,2]

[1]NLP & IR Group, Dpto. Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia (UNED), Madrid, 28040, Spain

[2]Instituto Mixto de Investigación - Escuela Nacional de Sanidad (IMIENS), Madrid, 28029, Spain

Abstract

This paper describes the participation of the LSI_UNED team in the first subtask of the BioNNE-R 2026 shared task, held within the BioASQ workshop at CLEF. The subtask focuses on biomedical relation extraction between nested named entities in English PubMed abstracts. We reformulate relation extraction as a text classification problem over entity pairs, augmented with a NO_RELATION label to explicitly model negative instances. Given the large number of candidate pairs and severe class imbalance, we apply a set of pruning strategies to reduce the search space and improve training efficiency. These constraints are based on entity-type compatibility, textual distance, and sentence boundaries. Our best system achieves a macro-F1 score of 0.4514 on the test set, corresponding to a 59.79% improvement over the baseline. Results show that biomedical domain-specific language models consistently outperform general-domain models, highlighting the importance of in-domain pretraining.

Keywords

Relation Extraction, Biomedical Natural Language Processing, Nested Named Entity Recognition, Imbalanced Data Classification, Domain-specific Pre-training

1. Introduction

This work presents the participation of our team, LSI_UNED, in the shared task BioNNE-R 2026 [1], developed under the 14th BioASQ Workshop [2] at the Conference and Labs of the Evaluation Forum (CLEF). The previous edition of the task, named BioNNE-L [3], focused on Named Entity Recognition (NER) in biomedical texts, particularly on the detection of nested entities. The 2026 edition of the task addresses the NLP challenge of relation extraction involving nested named entities, introducing additional challenges due to overlapping entity structures and the larger number of candidate entity pairs. There are three subtasks depending on the addressed languages: English, Russian and both. Our team has participated in the first one. The proposed system reformulates relation extraction as a text classification task over entity pairs with a NO_RELATION label for negative instances, with its primary contribution being a pruning setup based on entity-type compatibility, textual distance, and sentence boundaries to reduce the search space and improve training efficiency.

Relation extraction (RE) is a fundamental task in Natural Language Processing (NLP) that aims to identify and classify semantic relationships between entities in text. It plays a crucial role in transforming unstructured biomedical literature into structured knowledge, enabling applications such as knowledge graph construction, clinical decision support, and biomedical discovery. Recent work highlights the growing complexity of biomedical RE as the volume and diversity of biomedical publications increase [4]. This challenge is further exacerbated by the frequent occurrence of nested entities in biomedical texts, where one entity is fully or partially contained within another. Nested entities introduce ambiguity in entity boundaries and increase the number of candidate entity pairs, making it more difficult for RE models to correctly identify and classify relations [5, 6].

The remainder of this paper is organized as follows. Section 2 provides a brief overview of related work, as well as a description of the best-performing systems from the previous edition of the shared

CLEF 2026 Working Notes, 21 – 24 September 2026, Jena, Germany

*Corresponding author.

aramirez@lsi.uned.es (A. Ramirez-Arrabe); juanerd@lsi.uned.es (J. Martinez-Romo); aduque@lsi.uned.es (A. Duque)

0009-0001-8550-6275 (A. Ramirez-Arrabe); 0000-0002-6905-7051 (J. Martinez-Romo); 0000-0002-0619-8615 (A. Duque)

task. Section 3 describes the dataset and the proposed approach, detailing the models and tools employed. Section 4 presents the experimental results and performance analysis. Finally, Section 5 concludes the paper and discusses directions for future work.

2. Related work

In domains such as biomedicine, RE becomes particularly challenging due to the complexity of the terminology, the prevalence of long-distance dependencies, and the frequent occurrence of nested and overlapping entities [7]. Recent approaches have addressed these issues from different perspectives. For instance, the JITS framework [8] improves document-level RE by incorporating intermediate tasks that better capture cross-sentence semantics. Another study [7] proposes Bio-RFX, a model that predicts relations before performing entity extraction to reduce entity ambiguity.

The first edition of this challenge, named BioNNE-L [3], was focused on Named Entity Recognition (NER) in biomedical text. In particular, it addressed the detection of nested entities, which represent a major difficulty in clinical and scientific corpora due to their frequent overlap and hierarchical structure. The task considered three subtasks depending on the language (English, Russian, and multilingual) and three entity types: ANATOMY, DISO, and CHEM. The three best teams were verbanexialab [9], LYX_DMIIP_FDU [10], and BlancaPlanca [11]. Verbanexialab leveraged SapBERT [12] pretrained on UMLS concepts to obtain biomedical entity embeddings, combining cosine similarity with lexical and character-level matching signals. LYX_DMIIP_FDU fine-tuned a BERGAMOT-based model [13] using contrastive learning, enriching entity representations with contextual information from the surrounding text. BlancaPlanca also used BERGAMOT in a zero-shot retrieval setting based on embedding similarity, incorporating language-specific preprocessing such as Russian lemmatization. In the English subtask, verbanexialab and LYX_DMIIP_FDU shared first place with an F1-score of 0.74. In the Russian subtask, LYX_DMIIP_FDU and BlancaPlanca achieved the best results, with an F1-score of 0.76. Finally, in the multilingual subtask, LYX_DMIIP_FDU obtained 0.75 F1, followed by BlancaPlanca with 0.73. The 2026 edition of BioNNE shifts from an entity extraction task to a relation extraction task, which is generally considered more challenging. Therefore, lower F1-score values are expected.

3. Methodology

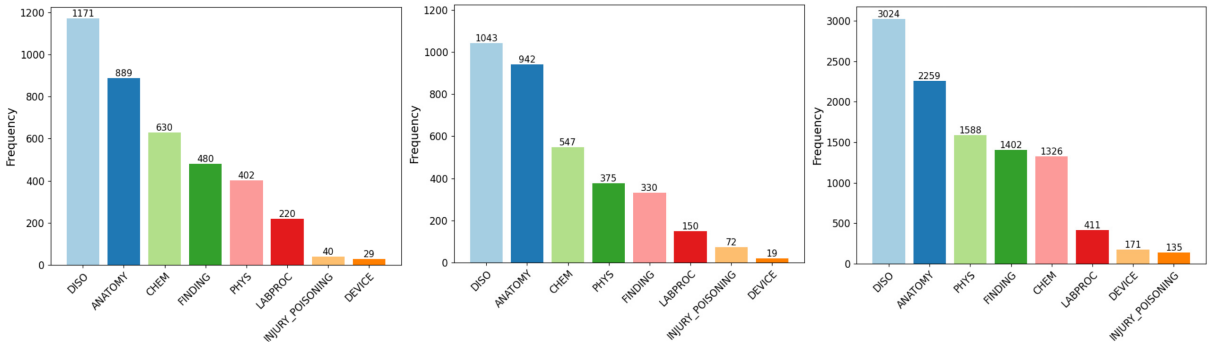
3.1. Dataset

The BioNNE-R task utilizes the NEREL-BIO dataset [14], a corpus of PubMed abstracts written in Russian and English that extends the NEREL corpus [15] by incorporating biomedical entity types. The corpus for the first subtask is entirely in English and contains 258 documents. It is split into a training set of 55 documents, a development set of 50 documents, and a test set of 153 documents. Table 1 presents the main statistics of the corpus. We can see that the number of entities is relatively high, with an average of approximately 67–70 entities per document.

Set	Training	Development	Test
Total documents	55	50	153
Total entities	3861	3478	10316
Mean entities per document	70.20	69.56	67.42
Relations	3193	2967	-

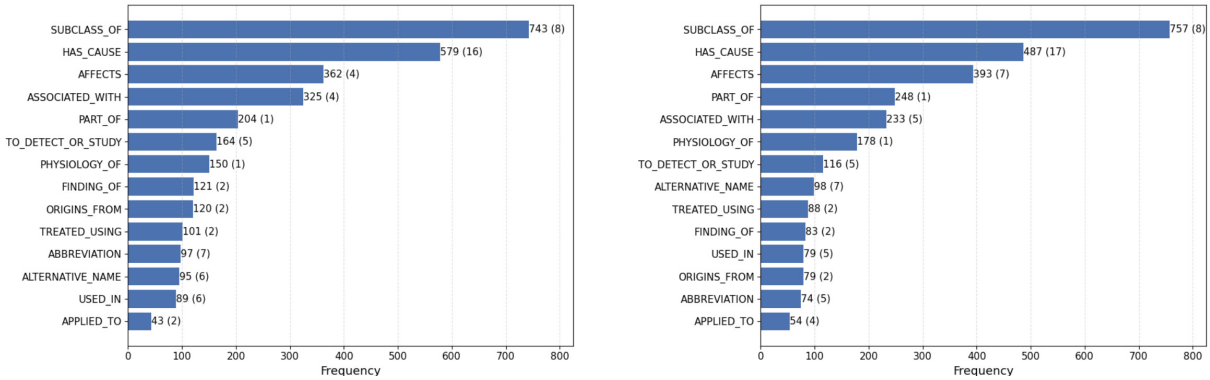
Table 1
Corpus statistics for Subtask 1 of the BioNNE-R 2026 corpus. The number of relations in the test set is not disclosed, as it is the target of prediction in the task.

The original corpus focuses on nested Named Entity Recognition (NER), meaning that many entities are nested or overlap with one another. It includes eight entity types: ANATOMY, CHEM, DEVICE, DISO, FINDING, INJURY_POISONING, LABPROC, and PHYS. Figure 1 shows the frequency distribution



(a) Training set (b) Development set (c) Test set

Figure 1: Frequency distribution of entity types across train, development and test sets.



(a) Training set (b) Development set

Figure 2: Frequency distribution of relation types across the training and development sets. The numbers in parentheses indicate the number of head–tail entity type combinations associated with each relation. The test set relations are not disclosed, as their prediction is the goal of the task.

of entity types across the three splits. As can be observed, the distribution is consistent across the sets. However, the entity types are highly imbalanced. DISO is the most frequent type, with thousands of occurrences, followed by ANATOMY. In contrast, DEVICE is the least frequent, with very few instances in each split, and INJURY_POISONING is the second least frequent category.

Entities in the corpus are paired to form relations. Entity types can form relations both with themselves and with other entity types. The relations are not bidirectional; that is, there is always a designated head entity and a tail entity. The corpus includes fourteen relation types: ABBREVIATION, AFFECTS, ALTERNATIVE_NAME, APPLIED_TO, ASSOCIATED_WITH, FINDING_OF, HAS_CAUSE, ORIGINS_FROM, PART_OF, PHYSIOLOGY_OF, SUBCLASS_OF, TO_DETECT_OR_STUDY, TREATED_USING, and USED_IN. Figure 2 shows the frequency distribution of relations across the training and development sets. The distribution across the splits is highly imbalanced. SUBCLASS_OF and HAS_CAUSE are the most frequent relation types, whereas APPLIED_TO is the least frequent, with fewer than 100 instances in total. In both plots, the numbers in parentheses indicate the number of valid entity-type combinations for each relation. For instance, PHYSIOLOGY_OF has only one valid combination, which is formed between PHYS and ANATOMY entities, in that order. In contrast, for example, HAS_CAUSE appears in the development set across 17 different combinations of entity types.

3.2. System description

The test set of the corpus includes the entity annotations, so we have gold-standard entities which, as shown in Table 1, are 10,316 in total. However, the relation annotations between entities are not provided, which is the goal of the task. For this reason, we introduced a new label called NO_RELATION,

which allows the model to predict the absence of a relation when given a pair of entities.

Therefore, we fine-tuned a text classification model that receives the following input for each pair of entities:

$$\texttt{[HEAD]}\ \text{entity}_A\ \texttt{[HTYPE]}\ \text{type}_A\ \texttt{[SEP]}\ \texttt{[TAIL]}\ \text{entity}_B\ \texttt{[TTYPE]}\ \text{type}_B$$

An example of its usage is shown below, considering two entities, *acute coronary syndrome* and *ACS*, both of type DISO and related by the ABBREVIATION relation:

```
[HEAD] acute coronary syndrome [HTYPE] DISO [SEP] [TAIL] ACS [TTYPE] DISO
```

3.2.1. Pruning criteria

The number of possible combinations of entity pairs considering all entities in each document is extremely large, especially given that there is an average of more than 67 entities per document across the different splits of the corpus. For instance, the number of possible combinations in the training set is 317,976, of which only 3,193 correspond to real relations (1.00%). As a result, each model takes hours to train on an extremely imbalanced dataset.

Therefore, by analyzing the corpus, we identified several characteristics that are common to most relations, and we defined three conditions that entity pairs must satisfy to be considered. An additional condition was later introduced after observing that some true relation pairs occurred very infrequently and introduced noise. Systems were evaluated both with and without this additional condition.

- **Condition 1:** Considering that there are eight entity types, and that relations can occur both within the same type and across different types, and are not bidirectional, there are 64 possible combinations of entity-type pairs. This condition restricts the pairs to only those entity-type combinations that appear as relations in either the training or development sets. In total, 37 such pairs are retained.
- **Condition 2:** Only pairs of entities whose distance within the text is less than 100 characters are included. We experimented with other thresholds (150 and 200 characters), but 100 yielded the best results.
- **Condition 3:** Only pairs of entities that belong to the same sentence are considered. Sentence segmentation was performed using the spaCy Python NLP library[1].
- **Extra condition:** It extends Condition 1 by restricting the pairs to those entity-type combinations that appear at least 50 times in total in the training and development sets. This further reduces the number of possible pairs to 23. We also experimented with other occurrence thresholds, particularly requiring combinations to appear at least 30 times, but a threshold of 50 yielded the best results.

Table 2 shows how the number of possible combinations is reduced after applying each condition. It can be observed that a certain proportion of true entity pairs is also discarded. However, this trade-off results in a more balanced corpus, as well as reduced computational complexity during training.

3.2.2. Pre-trained models

Several Transformer-based [16] pre-trained language models were fine-tuned and evaluated on the described text classification task. The selected models include both general-purpose and domain-specific architectures, as well as multilingual and monolingual variants. All models were fine-tuned using the same training configuration, with a maximum of 10 training epochs and early stopping with a patience of 3 epochs based on the macro-F1 score obtained on the development set, a batch size of 16, a learning rate of 2×10^{-5}, and a weight decay of 0.01.

[1] https://spacy.io/usage/linguistic-features

	Train set		Dev set	
	All possible combinations of entity pairs	True combinations of entity pairs	All possible combinations of entity pairs	True combinations of entity pairs
Total	317976	3193	287692	2967
Condition 1	239526	3193	215244	2967
+ Condition 2	27644	2830	26195	2712
+ Condition 3	21727	2806	20881	2682
+ Extra condition	19856	2743	19023	2597

Table 2
Reduction in the number of possible and true entity-pair combinations in the training and development sets after applying each pruning condition.

- bert-base-multilingual-cased[2] [17]: It has around 110M parameters and was pretrained on Wikipedia articles covering the top 104 languages.
- bert-base-uncased[3] [18]: It contains approximately 110M parameters and was pretrained on the BookCorpus dataset [19] and English Wikipedia.
- xlm-roberta-base[4] [20]: This model contains approximately 278M parameters and was trained on multilingual corpora derived from CommonCrawl, covering over 100 languages.
- biobert-v1.1[5] [21]: It has approximately 110M parameters and is initialized from BERT-base, further pre-trained on large biomedical corpora including PubMed abstracts and PubMed Central (PMC) full-text articles.
- Bio_ClinicalBERT[6] [22]: This model contains approximately 110M parameters, initialized from biobert-v1.1, and further pre-trained on clinical notes from the MIMIC-III database [23].

Among these models, bert-base-multilingual-cased and xlm-roberta-base are multilingual models, while the other three are monolingual English models. Regarding the domain, biobert-v1.1 and Bio_ClinicalBERT are biomedical-specific models, while the others are general-purpose models.

4. Experiments and results

The evaluation script provided by the organizers includes the metrics micro-Precision, micro-Recall, micro-F1 and macro-F1. Among them, the metric considered as the main one is macro-F1, which is the only one returned when evaluating the test set.

Table 3 reports the results obtained on the development set. As can be observed, systems with the extra condition consistently achieve higher micro-F1 and macro-F1 scores than those without it. The best-performing system in both F1 metrics is the one fine-tuned using the biobert-v1.1 model with the extra condition. We decided to submit all systems for the test evaluation, except for the two marked with *, corresponding to the model xlm-roberta-base, which obtained a macro-F1 score below the baseline of 0.3435. Despite being the largest model evaluated, its general-domain multilingual pretraining appears to be less effective for English biomedical relation extraction than the domain-specific pretraining of biobert-v1.1 and Bio_ClinicalBERT

The evaluation script also returns a classification report that provides a breakdown of the Precision, Recall, and F1-score for each relation type. Figure 3 presents the results of the best-performing system on the development set regarding the macro-F1 score. We can observe that the relations FINDING_OF and SUBCLASS_OF achieve the highest Precision, Recall and F1-score. They are followed by AFFECTS and ASSOCIATED_WITH. Notably, three of these relations are also among the four most frequent relation types in the dataset. In contrast, ALTERNATIVE_NAME appears to be the most challenging

[2]https://huggingface.co/google-bert/bert-base-multilingual-cased
[3]https://huggingface.co/google-bert/bert-base-uncased
[4]https://huggingface.co/FacebookAI/xlm-roberta-base
[5]https://huggingface.co/dmis-lab/biobert-v1.1
[6]https://huggingface.co/emilyalsentzer/Bio_ClinicalBERT

Model	Extra condition	Micro-Precision	Micro-Recall	Micro-F1	Macro-F1
bert-base-multilingual-cased	No	0.4178	0.4738	0.4440	0.3468
bert-base-multilingual-cased	Yes	**0.4704**	0.4596	0.4649	0.3709
bert-base-uncased	No	0.3817	0.5066	0.4353	0.3452
bert-base-uncased	Yes	0.4333	0.4873	0.4587	0.3635
xlm-roberta-base	No	0.4019	0.4809	0.4379	0.3171*
xlm-roberta-base	Yes	0.4575	0.4664	0.4619	0.3399*
biobert-v1.1	No	0.4104	**0.5404**	0.4665	0.3852
biobert-v1.1	Yes	0.4575	0.5245	**0.4887**	**0.4013**
Bio_ClinicalBERT	No	0.3863	0.5025	0.4368	0.3509
Bio_ClinicalBERT	Yes	0.4312	0.4863	0.4571	0.3676

Table 3
Performance results for each system on the development set. The best result for each metric is highlighted in bold, while the second-best result is underlined. The baseline system provided by the organizers achieved an macro-F1 score of 0.3435. * Results correspond to systems whose macro-F1 score falls below the baseline and were not included in the final submissions.

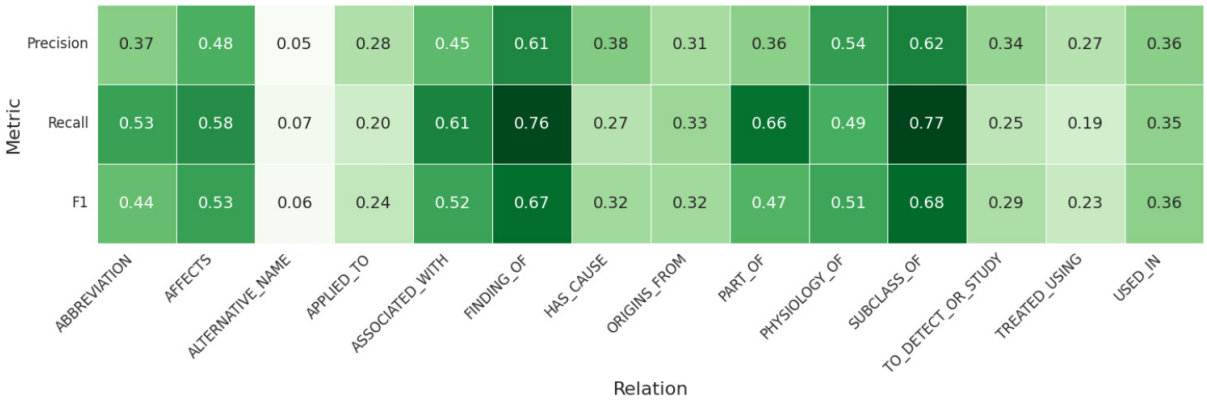


Figure 3: Heatmap of the classification report, organized by relation type, for the best macro-F1 system on the development set. The system uses the biobert-v1.1 model with the extra condition.

relation type, with very few instances correctly identified. This type of relation accounts for 3.13% of the total relations in the training and development sets. One possible explanation is that, as shown in Figure 2, although ALTERNATIVE_NAME is among the least frequent relation types, it is associated with a relatively large number of different entity-type pairs. Consequently, its limited training instances are spread across a more diverse set of patterns, making the relation more difficult for the models to learn.

To fine-tune the models for the test evaluation, we employed both the training and development data in order to increase the number of instances available during training. Table 4 shows the results achieved by the submitted systems. In this case, only the macro-F1 score was provided by the evaluation platform. The best macro-F1 score, 0.4514, was obtained by the system based on the monolingual biomedical pre-trained model biobert-v1.1 with the extra condition applied. This result ranked twelfth out of 23 submissions, while the winning team achieved a macro-F1 score of 0.506. The baseline system provided by the task organizers achieved a macro-F1 score of 0.2825, meaning that our best system represents an improvement of 59.79%. Our second-best system, also employing the extra condition, was based on the biomedical pre-trained model Bio_ClinicalBERT and achieved a macro-F1 score of 0.4317.

5. Conclusions

This work presents the participation of the LSI_UNED team in the BioNNE-R shared task on biomedical relation extraction using the NEREL-BIO corpus. The task focuses on identifying relations between nested biomedical entities in English PubMed abstracts. We addressed the task by formulating relation extraction as a text classification problem over pairs of entities and introduced a new NO_RELATION

Model	Extra condition	Macro-F1
biobert-v1.1	Yes	**0.4514**
Bio_ClinicalBERT	Yes	<u>0.4317</u>
biobert-v1.1	No	0.4276
bert-base-uncased	Yes	0.4176
bert-base-multilingual-cased	Yes	0.4121
Bio_ClinicalBERT	No	0.4061
bert-base-uncased	No	0.4001
bert-base-multilingual-cased	No	0.3823

Table 4
Performance results of the submitted systems on the test set, sorted by macro-F1 in descending order. The best result is highlighted in bold, while the second-best is underlined. The baseline system provided by the organizers achieved a macro-F1 score of 0.2825.

label to model the absence of relations. Given the large number of possible entity pairs and the severe class imbalance of the corpus, we proposed several pruning criteria based on entity-type compatibility, textual distance, and sentence boundaries. In addition, we introduced an extra filtering condition that removes infrequent entity-type combinations, which proved beneficial for reducing noise and improving overall performance.

The experimental results showed that the systems that incorporate the extra condition consistently outperformed those without it. The best-performing system achieved a macro-F1 score of 0.4514 on the test set. Furthermore, the results indicate that domain-specific biomedical pre-training plays a key role in relation extraction performance, as biomedical language models consistently obtained the highest scores.

Despite the improvements obtained, the main limitation that the proposed approach presents is that the pruning strategies, although effective for reducing computational complexity and dataset imbalance, discard a proportion of true relations, which may negatively affect Recall.

Future work will focus on several directions. First, it would be interesting to explore additional strategies to reduce the number of candidate entity pairs while minimizing the loss of true relations. Along this line, it would also be worthwhile to investigate class weighting, downsampling, or other imbalance mitigation techniques to address the class imbalance between the NO_RELATION label and the true relation labels. Another promising direction would be the use of more recent architectures, including Large Language Models (LLMs), to assess their capability to model complex biomedical relations.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Science and Innovation within the EDHER-MED Project under grant PID2022-136522OB-C21 as well as by the Universidad Nacional de Educación a Distancia (UNED), Spain within project SICAMESP (2023-VICE-0029).

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to: Grammar and spelling check.

References

[1] I. Rozhkov, N. Loukachevitch, E. Tutubalina, Overview of the BioASQ BioNNE-R Task on Nested Biomedical Relation Extraction at CLEF 2026, in: E. Sánchez Salido, A. Barrón-Cedeño, A. García Seco de Herrera, S. MacAvaney, J. M. Struß (Eds.), CLEF 2026 Working Notes, 2026.

[2] A. Nentidis, G. Katsimpras, A. Krithara, M. Krallinger, M. Rodríguez-Ortega, E. Rodriguez-López, N. Loukachevitch, I. Rozhkov, E. Tutubalina, D. Dimitriadis, V. Patsiou, G. Tsoumakas, G. Giannakoulas, A. Bekiaridou, A. Samaras, G. M. Di Nunzio, N. Ferro, S. Marchesin, M. Martinelli, G. Silvello, G. Paliouras, Overview of BioASQ 2026: The fourteenth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering, volume Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Seventeenth International Conference of the CLEF Association (CLEF 2026) of *Lecture Notes in Computer Science*, Springer, 2026.

[3] A. Sakhovskiy, N. V. Loukachevitch, E. Tutubalina, Overview of the bioasq bionne-l task on biomedical nested entity linking in CLEF 2025, in: Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2025, Madrid, Spain, 9-12 September 2025, CEUR Workshop Proceedings, CEUR-WS.org, 2025, pp. 41–50. URL: https://ceur-ws.org/Vol-4038/paper_3.pdf.

[4] Z. Zhang, M. Fang, R. Wu, H. Zong, H. Huang, Y. Tong, Y. Xie, S. Cheng, Z. Wei, M. J. C. Crabbe, et al., Large-scale biomedical relation extraction across diverse relation types: model development and usability study on covid-19, Journal of Medical Internet Research 25 (2023) e48115.

[5] M. Ju, M. Miwa, S. Ananiadou, A neural layered model for nested named entity recognition, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018, pp. 1446–1459.

[6] Y. Wang, H. Tong, Z. Zhu, Y. Li, Nested named entity recognition: a survey, ACM Transactions on Knowledge Discovery from Data (TKDD) 16 (2022) 1–29.

[7] M. Wang, F. Liu, X. Li, B. Dong, Z. Li, T. Pan, J. Wang, Bio-rfx: Refining biomedical extraction via advanced relation classification and structural constraints, in: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024, Association for Computational Linguistics, 2024, pp. 10524–10539. URL: https://doi.org/10.18653/v1/2024.emnlp-main.588.

[8] J. Li, D. Pan, Z. Yang, Y. Sun, H. Lin, J. Wang, JTIS: enhancing biomedical document-level relation extraction through joint training with intermediate steps, Database J. Biol. Databases Curation 2024 (2024). URL: https://doi.org/10.1093/database/baae125.

[9] D. P. Gnecco, J. Serrano, E. Puertas, J. C. M. Santos, Hybrid re-ranking for biomedical entity linking using sapbert embeddings: A high-performance system for bionne-l 2025-1, in: Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2025, Madrid, Spain, 9-12 September 2025, CEUR Workshop Proceedings, CEUR-WS.org, 2025, pp. 497–508. URL: https://ceur-ws.org/Vol-4038/paper_35.pdf.

[10] Y. Liu, S. Zhu, Lyx_dmiip_fdu at bioasq 2025: Utilizing bert embeddings for biomedical text mining, in: CLEF, 2025.

[11] A. Burlova, Navigating partial umls terminology: Gat embeddings and confidence analysis for multilingual concept linking, in: CLEF, 2025.

[12] F. Liu, E. Shareghi, Z. Meng, M. Basaldella, N. Collier, Self-alignment pretraining for biomedical entity representations, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021, Association for Computational Linguistics, 2021, pp. 4228–4238. URL: https://doi.org/10.18653/v1/2021.naacl-main.334.

[13] A. Sakhovskiy, N. Semenova, A. Kadurin, E. Tutubalina, Biomedical entity representation with graph-augmented multi-objective transformer, in: Findings of the Association for Computational Linguistics: NAACL 2024, 2024, pp. 4626–4643.

[14] N. V. Loukachevitch, S. Manandhar, E. Baral, I. Rozhkov, P. Braslavski, V. Ivanov, T. Batura, E. Tutubalina, NEREL-BIO: a dataset of biomedical abstracts annotated with nested named entities, Bioinform. 39 (2023). URL: https://doi.org/10.1093/bioinformatics/btad161.

[15] N. V. Loukachevitch, E. Artemova, T. Batura, P. Braslavski, I. Denisov, V. Ivanov, S. Manandhar, A. Pugachev, E. Tutubalina, NEREL: A russian dataset with nested named entities, relations and events, in: Proceedings of the International Conference on Recent Advances in Natural Language

Processing (RANLP 2021), Held Online, 1-3 September, 2021, INCOMA Ltd., 2021, pp. 876–885. URL: https://aclanthology.org/2021.ranlp-1.100.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[17] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), Association for Computational Linguistics, 2019, pp. 4171–4186. URL: https://doi.org/10.18653/v1/n19-1423.

[18] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805.

[19] Y. Zhu, R. Kiros, R. S. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, S. Fidler, Aligning books and movies: Towards story-like visual explanations by watching movies and reading books, in: 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015, IEEE Computer Society, 2015, pp. 19–27. URL: https://doi.org/10.1109/ICCV.2015.11.

[20] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, Association for Computational Linguistics, 2020, pp. 8440–8451. URL: https://doi.org/10.18653/v1/2020.acl-main.747.

[21] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, Biobert: a pre-trained biomedical language representation model for biomedical text mining, Bioinform. 36 (2020) 1234–1240. URL: https://doi.org/10.1093/bioinformatics/btz682. doi:10.1093/BIOINFORMATICS/BTZ682.

[22] E. Alsentzer, J. R. Murphy, W. Boag, W. Weng, D. Jin, T. Naumann, M. B. A. McDermott, Publicly available clinical BERT embeddings, CoRR abs/1904.03323 (2019). URL: http://arxiv.org/abs/1904.03323.

[23] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R. G. Mark, Mimic-iii, a freely accessible critical care database, Scientific data 3 (2016) 1–9.