Overview of the BioASQ BioNNE-L Task on Biomedical Nested Entity Linking in CLEF 2025

Andrey Sakhovskiy[1,2,3], Natalia Loukachevitch[4,*] and Elena Tutubalina[1,5,*]

[1]Kazan Federal University, Russia
[2]Sber AI, Russia
[3]Skoltech, Russia
[4]Lomonosov Moscow State University, Russia
[5]Artificial Intelligence Research Institute, Russia

Abstract

The task of biomedical entity linking (EL), which is intended to normalize a free-form textual entity to a concept from a standardized domain-specific vocabulary, is foundational for factuality-sensitive applications. Despite vast research on EL, modern methods ignore the nested structure of longer entities, which may provide vital context for joint normalization of nested entities. This paper presents an official results report for the **BioNNE-L**, a shared task on **B**iomedical **N**ested **N**amed **E**ntity **L**inking conducted within BioASQ 2025 Workshop on biomedical semantic indexing and question answering. The shared task included three subtasks organized into two evaluation tracks: monolingual track with (i) English and (ii) Russian subtasks, and (iii) multilingual track combining the data from the two monolingual subtasks. For evaluation, two novel test sets of annotated entities are released, each containing 154 PubMed abstracts in English and Russian. The evaluation of system submissions from 7 participating teams has revealed the effectiveness of small domain-specific models for nested entity linking even in the era of large language models.

Keywords

BioNLP, Biomedical NLP, Nested Entity Linking, Biomedical Text Mining, Domain-specific Language Models

1. Introduction

Recent progress in nested Named Entity Recognition (NER) has led to the creation of richly annotated datasets [1, 2] such as NEREL [3] and NEREL-BIO [4]. A recent dataset, NEREL [3], is annotated with over 56k named entities of 29 types, while its biomedical extension, NEREL-BIO [4], is annotated with over 70k entities of 37 types. However, entity linking has not yet been extended to address nested entities. Most of the existing work on biomedical EL continues to focus on a non-nested formulation despite the high nestedness of biomedical entities [5, 6, 7, 8, 9]. The recently released entity linking annotation for nested entities from the NEREL-BIO corpus [10] as well as the annotation guidelines enable research on joint normalization for nested entities.

Inspired by the success of BioNNE [11], a shared task on nested NER held within the BioASQ 2024 workshop [12], we extend the research of nested entities to entity linking. This paper provides a detailed overview of the Biomedical Nested Named Entity Linking (BioNNE-L) which is focused on exploring nested EL formulation by using annotated PubMed abstracts and is part of the BioASQ 2025 workshop. We setup an evaluation pipeline for entity linking over NEREL-BIO entities as well as newly annotated entities from BioNNE data.

All BioNNE-L materials can be found on the shared task's GitHub[1] and Codalab pages[2]. Annotated data and normalization dictionary are also available at HuggingFace[3].

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain
*Corresponding author.
andrey.sakhovskiy@gmail.com (A. Sakhovskiy); louk_nat@mail.ru (N. Loukachevitch); tutubalinaev@gmail.com (E. Tutubalina)
0000-0003-2762-2910 (A. Sakhovskiy); 0000-0002-1883-4121 (N. Loukachevitch); 0000-0001-7936-0284 (E. Tutubalina)

[1]https://github.com/nerel-ds/NEREL-BIO/tree/master/BioNNE-L_Shared_Task
[2]https://codalab.lisn.upsaclay.fr/competitions/21568
[3]https://huggingface.co/datasets/andorei/BioNNE-L

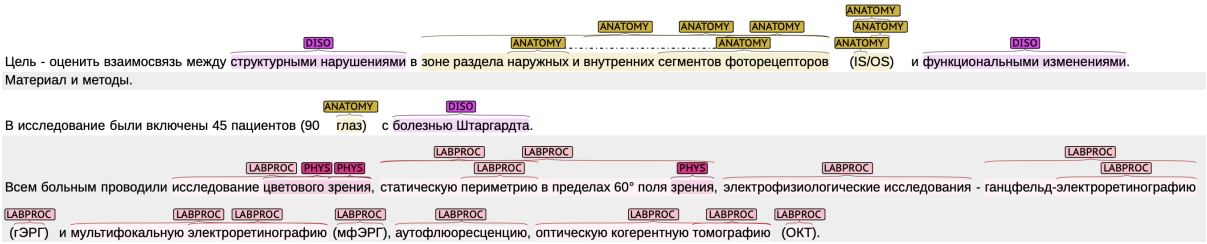

Figure 1: Sample of annotations of nested entities in the Russian abstract (PMID: 27456564).

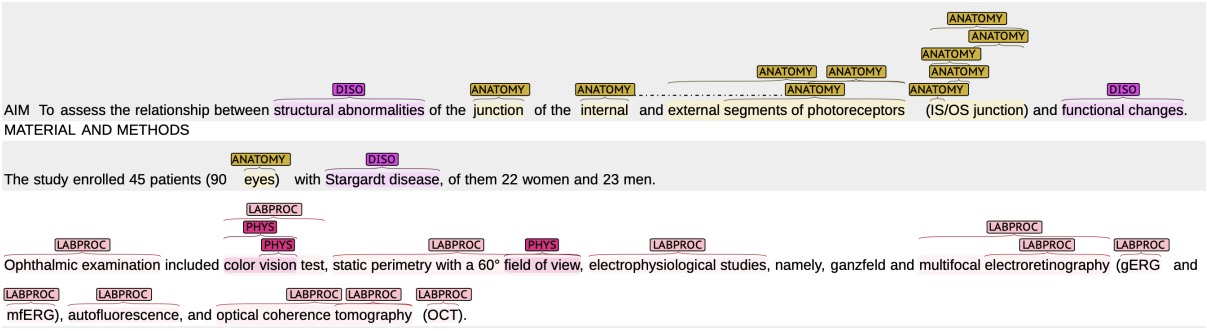

Figure 2: Sample of annotations of nested entities in the English abstract (PMID: 27456564).

2. BioNNE-L Shared Task

In the BioNNE-L Shared Task, we address the medical entity linking task, also known as Medical Concept Normalization (MCN), which is to map given entities to the most relevant vocabular entries from an external source, e.g., concepts from the UMLS metathesaurus [13] identified with concept unique identifiers (CUIs). Although the task has been widely explored in recent years, existing approaches usually treat each entity individually, medical entities often form a nested structure, where an entity can be a subpart of another entity. One of the key features of BioNNE-L is the focus on nested entities that are (i) derived from the MCN annotation of the NEREL-BIO corpus [4, 10] and (ii) supplemented by newly annotated data in both English and Russian. The annotated entity types are disorders (DISO), anatomical structures (ANAT), and chemicals (CHEM). The competition was organized into three subtasks that fell under two evaluation tracks:

1. **Monolingual track** that treated English and Russian data independently;
2. **Bilingual track** that required a single bilingual model for the combined Russian and English data.

3. Dataset

Training and validation sets for the BioNNE-L competition are based on the NEREL-BIO dataset [4] and additional annotated texts for the BioNNE competition organized in 2024 [11]. NEREL-BIO is a corpus of PubMed abstracts written in Russian and English. It enhances the NEREL [3] dataset, originally designed for the general domain, by incorporating biomedical entity types. Biomedical entity types in NEREL-BIO are annotated according to UMLS definitions of relevant concepts. All the abstracts are annotated in the BRAT format [14].

Figures 1 and 2 present parallel examples of nested named entities in NEREL-BIO for one abstract. Table 1 provides a comprehensive list of entity types, along with their explanations and examples.

Compared to the original NEREL-BIO and BioNNE datasets, we selected only three most common entity types for the BioNNE-L competition: disorders (*DISO*), anatomical structures (*ANAT*), and chemicals (*CHEM*).

Table 1
List of entity types provided in the BioNNE-L Shared Task.

Entity type	Specification	Examples
DISO	any deviations from normal state of organ- ism: diseases, symptoms, abnormality of organ, excluding injuries or poisoning	lumbar vertebral canal stenosis, exogenous allergic alveolitis, appendicitis, haemorrhoids, magnesium deficiency dysfunctions, arteriovenous angiodysplasia, type 2 diabetes mellitus
CHEM	chemicals including legal and illegal drugs, biological molecules	venlafaxine, resistin, lipoprotein, mydocalm-richter, leptin, melatonin, opioid, iodine, adrenalin, isotonic NaCl solution
ANAT	organs, body parts, cells and cell components	epidermal nerve fibers, skin biopsy specimens, tumor tissue, chiasmatic-sellar area, blood, low back, eye, bone, brain, lower limb, oral cavity

Table 2
BioNNE-L 2025 statistics for Disorder (**DISO**), Chemical (**CHEM**), and Anatomical Structure (**ANAT**) among Russian and English entities as well as normalization dictionary statistics.

Entity type	Refined NEREL-BIO				Novel data		Dictionary	
	Train		Dev		Test			
	Ru	En	Ru	En	Ru	En	Ru	En
# documents	716	54	50	50	154	154	—	—
Number of entities								
DISO	11,168	1,200	925	1,029	2,811	3,068	91,867	1,825,048
CHEM	4,741	579	531	564	1,218	1,345	47037	1,732,096
ANAT	8,346	911	878	901	2,186	2,248	6899	345,043
	24,255	2,690	2,334	2,494	6,215	6,661	145,803	3,902,187
Number of unique UMLS CUIs								
DISO	2,001	489	374	397	770	879	49,358	641,273
CHEM	1,107	257	222	237	392	418	10,201	715,241
ANAT	1,146	376	278	297	548	598	1,806	153,917

The resulting dataset comprises 662 annotated PubMed abstracts in Russian and 104 parallel abstracts in Russian and English. 104 parallel abstracts were randomly split for training and validation sets for each subtask. A novel test set was developed for the shared task, consisting of 154 abstracts in English and Russian. Russian and English texts in dev and test sets are parallel texts written by the authors. The Russian training set contains Russian variants of English training texts. When annotating UMLS links, annotators worked with both Russian and English parallel texts and labeled the same entities and created the same links, if possible.

Table 2 shows the number of entities represented in each part of the data set. Observations can be summarized as follows. First, entities labeled as DISO and ANAT are the most frequent across all sets, with DISO being particularly prevalent in both training and test sets. Second, it can be seen that the numbers of entities in the English test set (EN test) and the Russian test set (RU test) are relatively comparable: the overall difference in numbers is about 7%. The English test set also shows a slightly higher number of unique CUIs, e.g. for DISO, 879 (En) vs. 770 (Ru). This may indicate that the English test data might pose a greater normalization challenge. Third, the size of the normalization dictionary reflects the much greater maturity and coverage of UMLS for English biomedical texts. The size difference in dictionary coverage suggests that English normalization benefit from broader recall, while Russian normalization faces potential coverage gaps and might need domain-specific expansions. The ANAT dictionary is much smaller than DISO or CHEM for both languages. This likely reflects a more finite set of anatomical terms compared to the expansive terminologies for diseases and chemicals.

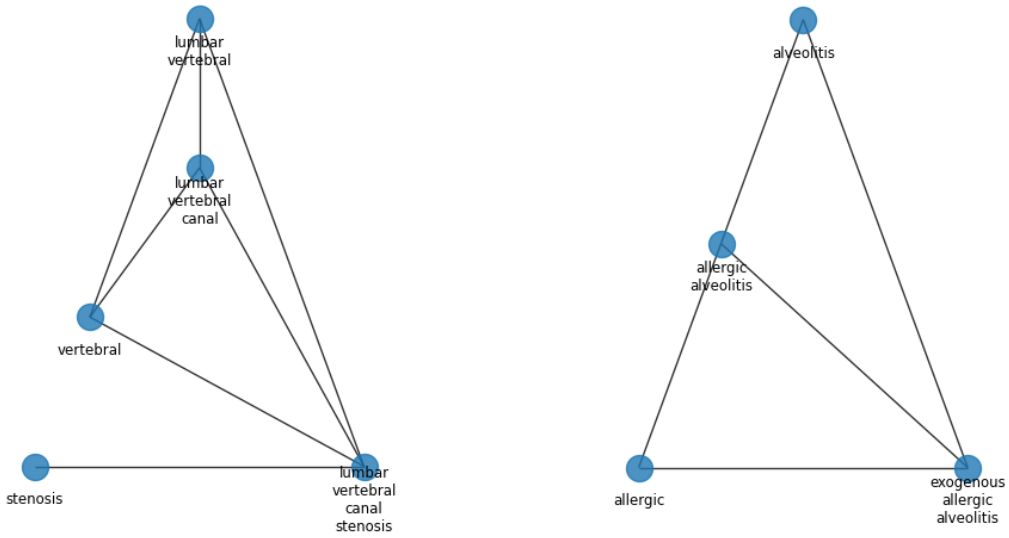

Figure 3: Examples of nested entities from the BioNNE-L English test set visualized as graphs. An edge connects two entity nodes when one entity is nested into another.

Normalization Dictionary As a normalization dictionary, we collect the English and Russian UMLS concepts filtered by the concept types *DISO*, *CHEM*, and *ANAT*. In UMLS, each concept is identified with a Concept Unique Identified (CUI) and a set of concept names in different languages including English and Russian.

3.1. BioNNE-L Challenges

The two key challenges of BioNNE-L data are inherited from NEREL-BIO [10]. First, the data exhibit a high level of nestedness, i.e., cases where a shorter entity is a subpart of a longer one. Some illustrative examples of nested entities are presented in Figure 3. The research question of whether nested entities would be linked more effectively when addressed jointly rather than individually remains unexplored. The second challenge is the incompleteness of vocabular terminology in a target low-resource language, e.g., Russian. NEREL-BIO's data annotation protocol addresses the issue by linking the entities absent in Russian UMLS to a concept that only has an English name. Although being well-aligned with the real-world terminology incompleteness scenario, cross-lingual annotation causes extensive normalization dictionary growth.

3.1.1. Problems of Annotating Texts with UMLS concepts

There are some problems in annotating Russian and English texts with UMLS concepts.

1) Lack of Russian translations in UMLS. When linking Russian texts with UMLS, there is a serious problem of the absence of many Russian terms in UMLS concept variants: The Russian part of UMLS includes only Russian translations for 1. 96% of the English UMLS concepts [10]. In such cases, the annotators tried to identify an appropriate UMLS concept even when a Russian translation was not available, which could require significant effort. For example, Russian translations of well-known single-word terms such as Complication (C0009566), Cirrhosis (C0023890), Bone (C0023890) are currently absent in UMLS.

In difficult cases, direct translation of a Russian medical term does not give a correct English term. To find a correct link to the UMLS concept, annotators should search for an appropriate English translation using various sources of information, including Latin terms for anatomical structures, Wikipedia pages in Russian and English, and even Russian scientific papers with English abstract and keywords [10].

2) Difficulties with assigning adjectives in UMLS. In domain-specific texts, adjectives can express some concepts. Therefore, in our detailed, nested annotation, adjectives should also be linked to UMLS concepts. However, in some cases, there can be two different concepts for nouns and adjectives with the same denotation, such as lung (C0024109) and pulmonary (C2709248). For veins, there are three relevant concepts: noun **vein** is mentioned in the C0042449 concept, and adjective *venous* is appeared in two concepts: C0042449 and C0348013 in UMLS. In some other cases, concept-related adjectives are absent in UMLS. For example, none of the Russian or English adjectives for the noun *nitrogen* (*nitric, nitrous*) are included in UMLS.

3) Ambiguity of terms in UMLS. Some non-ambiguous medical terms are assigned to several concepts in UMLS (see also [8]). For example, the term *cognitive disorders* is mentioned in C0009241 (Cognition disorders) and C0338656 (Impaired cognition) concepts. The term *thrombin* is assigned as a synonym to both concepts Thrombin (C0040018) and Thrombin test (C0863178), et al.

All the above-mentioned difficulties of manual linking to UMLS concepts also cause problems in automatic linking.

3.1.2. Differences in Annotations of English and Russian Texts

As we mainly dealt with English and Russian parallel texts, we can describe sources of different annotations in English and Russian.

1) The same concept is expressed in one language by a single word and in another language by a phrase. For example, the term *blood flow* is expressed as a single word in Russian, which means that in English texts two nested entities (*blood* and *blood flow*) are annotated, but only a single entity is linked to UMLS in Russian. Single-word term *brain* has corresponding two-word Russian phrase (головной мозг).

2) A single word in one language corresponds to a multi-word term in another language. For example, English term *reductase* exists only as a root in Russian. This leads to annotation of three entities and links in English for the term *Glutathione Reductase* and only a single link for its Russian translation (Глутатиоредуктаза).

3) Differences in syntactic structure and word order across languages lead to variations in the nestedness of multi-word terms. For example, in English term *Lower limb deep vein* the following UMLS concepts were revealed: vein (C0042449), deep vein (C0226514), limb (C0015385), lower limb (C0023216). In the Russian translation, additional Russian term (глубокие вены нижних конечностей – deep veins of lower limbs) and additional concept C0226813 (Structure of vein of lower extremity) can also be identified and linked to the corresponding UMLS concept.

4. Experiments

4.1. Evaluation metrics

Following prior research on entity linking [10, 6, 9, 5], we address BioNNE-L as a retrieval task: given a mention, a model must retrieve the top-k concepts from the given UMLS dictionary and employ two ranking-based evaluation metrics: (i) Accuracy@k and (ii) Mean Reciprocal Rank (MRR). Accuracy@k: Accuracy@k=1 if the correct UMLS CUI is retrieved at rank $\leq k$, and Accuracy@k=0 otherwise. $MRR = \frac{1}{|E|}\sum_{e \in E}\frac{1}{rank_e}$, where E is the set of entities, $|E|$ is the number of entities, $rank_e$ is the rank of entity e's the first correctly retrieved concept among the top k retrieved concepts.

4.2. Baseline

As a baseline, we adopt zero-shot ranking with each entity type processed independently to reduce the memory footprint caused by an extensive dictionary. Both input entities and normalization dictionary concepts are encoded with a BERGAMOT [5]. BERGAMOT adopts the power of BERT [15] and graph neural networks to capture both inter-concept and intra-concept interactions from the multilingual

Table 3
Overview of the approaches presented by participants for the BioNNE task. EN stands for the English-oriented and RU for the Russian-oriented tracks.

Team	Track	Approach
verbanexialab	EN	SapBERT w/ lexical and semantic reranking
LYX_DMIIP_FDU	Bilingual,EN,RU	BERGAMOT fine-tuning
BlancaPlanca	Bilingual,EN,RU	BERGAMOT w/ language-specific preprocessing
MSM Lab	Bilingual,EN,RU	SapBERT, BiomedBERT, two-step retrieval and ranking pipeline
dstepakov	Bilingual,RU	RoBERTa fine-tuning with contrastive learning
ICUE	Bilingual,EN,RU	BERT, BioSyn, LLM 0-shot reranking
NLPIMP	Bilingual	Russian LaBSE model pre-trained on medical data

UMLS graph. This model utilizes contrastive loss on textual and graph concept representations from UMLS to make them less sensitive to surface forms and enable intermodal knowledge exchange. For each entity, we rank all dictionary entries based on their dot product with the entity's embedding obtained from the BERGAMOT checkpoint[4] with *[CLS]* pooling. Finally, dictionary entries with the highest scores are retrieved as matching UMLS concepts.

4.3. Official Results

In total, we've received 23 Codalab registrations for the BioNNE-L task, with 7 teams submitting predictions during the evaluation phase. The systems submitted by the participants are summarized in Table 3. Most of the participants reported systems based on domain-specific biomedical BERT models [15], such as SapBERT [6], BERGAMOT [5], BiomedBERT [16].

Team **verbanexialab** [17] leveraged a SapBERT [6], pre-trained on UMLS concepts, to obtain entity embeddings, followed by a multicomponent re-ranking. They combined embedding cosine similarity with Jaccard similarity for lexical overlap recognition and Levenshtein distance for character-level alignment.

Team **LYX_DMIIP_FDU** [18] fine-tuned a BERGAMOT [5] model for each task via contrastive learning using the train- and dev-set entities to enrich the original vocabularies. The textual context of each entity was used as additional input to enhance the entity representation.

Team **BlancaPlanca** [19] used BERGAMOT for zero-shot retrieval based on entity-concept cosine similarity. They apply language-specific lemmatization for Russian and speed up the inference by chucking the normalization dictionary into type-specific parts of 100k entries each.

Team **MSM Lab** [20] adopted two-step retrieval and ranking pipeline. For English, they employ English SapBERT[5] [6] and BioMedBERT[6] [16] as retrieval and ranking models, respectively. For Russian and multilingual subtasks, they use multilingual SapBERT[7] [21] for both components.

Team **dstepakov** performed the nearest-neighbor search based on the cosine similarity of RoBERTa embeddings [22], fine-tuned contrastively on anchor-positive-negative term triplets via the InfoNCE objective [23].

Team **ICUE** [24] fine-tuned BioSyn [25] using the vocabularies reduced to less than 100k entries each. They fine-tune a separate BERT-based model [26] for English [27], Russian[8], and multilingual [28] subtasks, respectively. They re-ranked the initial retrieval results with *DeepSeek-R1-Distill-Llama-8B*[9].

Team **NLPIMP** performed the zero-shot ranking using a Russian LaBSE [29] model[10] pre-trained contrastively on an in-house Russian medical corpus.

[4]https://huggingface.co/andorei/BERGAMOT-multilingual-GAT
[5]https://huggingface.co/cambridgeltl/SapBERT-from-PubMedBERT-fulltext
[6]https://huggingface.co/microsoft/BiomedNLP-BiomedBERT-base-uncased-abstract-fulltext
[7]https://huggingface.co/cambridgeltl/SapBERT-UMLS-2020AB-all-lang-from-XLMR-large
[8]https://huggingface.co/KoichiYasuoka/bert-base-russian-upos
[9]https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Llama-8B
[10]https://huggingface.co/sergeyzh/LaBSE-ru-turbo

Table 4
Official evaluation results of the BioNNE-L task for the multilingual and monolingual tracks in terms of Accuracy@1 (**@1**), Accuracy@5 (**@5**), and MRR. The best results for each subtask and metric are highlighted in **bold**.

Team	Multilingual				English				Russian			
	#	@1	@5	MRR	#	@1	@5	MRR	#	@1	@5	MRR
verbanexialab [17]	—	—	—	—	1	**0.70**	0.80	**0.74**	—	—	—	—
LYX_DMIIP_FDU [18]	1	**0.68**	**0.84**	**0.75**	2	0.66	**0.84**	**0.74**	2	0.71	**0.84**	**0.76**
BlancaPlanca [19]	2	0.67	0.81	0.73	3	0.64	0.83	0.72	1	**0.72**	0.83	**0.76**
MSM Lab [20]	3	0.63	0.76	0.69	4	0.64	0.82	0.71	4	0.65	0.74	0.69
dstepakov	4	0.63	0.71	0.66	—	—	—	—	3	0.70	0.76	0.72
ICUE [24]	5	0.58	0.76	0.66	6	0.51	0.79	0.62	5	0.62	0.72	0.67
baseline	6	0.53	0.70	0.60	5	0.57	0.78	0.66	6	0.52	0.59	0.55
NLPIMP	7	0.41	0.58	0.48	—	—	—	—	—	—	—	—

The official evaluation results, ordered by Accuracy@1 value, for BioNNE-L are summarized in Table 4. Team LYX_DMIIP_FDU ranked first in the multilingual track and second in the two monolingual subtasks by fine-tuning BERGAMOT. Top 1 results for the Russian and English data are achieved by multilingual BERGAMOT (Team BlancaPlanca) and English SapBERT (Team verbanexialab) models, respectively. Despite using LLM-based re-ranking, Team ICUE did not surpass BERT-only systems.

Overall, the results show that performance on the multilingual track is consistently lower than on the monolingual English or Russian tracks. Most teams have a drop of about 5 to 10 percentage points in Accuracy@1 when moving from monolingual to multilingual settings. This may highlights that cross-lingual biomedical entity normalization remains more challenging than working within a single language, likely due to differences in terminology, translation ambiguities, and vocabulary coverage.

5. Conclusion

This paper presents an overview of the official evaluation results for the BioNNE-L shared task on biomedical nested entity linking. The evaluation was organized into three tracks: English, Russian, and bilingual, and aimed at normalization of disorders, chemicals, and anatomical structure mentions to the UMLS vocabulary. The best results were achieved by BERT-based normalization approaches. Top-performing systems for bilingual and Russian tracks adopted multilingual BERGAMOT which is a BERT model pre-trained on textual and graph data from the UMLS metathesaurus. The best English system re-ranked SapBERT's retrieval results through lexical and character-level similarity scores. In general, the evaluation results have proven the effectiveness of compact domain-specific encoders for nested entity linking.

Future work should focus on addressing the critical gaps identified in this shared task. This includes expanding cross-lingual terminology for Russian UMLS by utilizing semi-automated pipelines that leverage machine translation of English UMLS entries, which can be validated by human experts or LLMs. Additionally, mining Russian clinical literature and utilizing resources like Wikidata will enhance this process. Furthermore, employing joint modeling of nested entity hierarchies through graph-based architectures, such as Graph Neural Networks (GNNs), could help propagate contextual constraints between parent and child entities, thereby resolving ambiguities.

Acknowledgments

This work has been supported by the Russian Science Foundation grant # 23-11-00358. We would like to thank all the participating teams who contributed to the success of the shared task through their interesting approaches and experiments.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check and Improve writing style. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] H. Ming, J. Yang, L. Jiang, Y. Pan, N. An, Few-shot nested named entity recognition, arXiv preprint arXiv:2212.00968 (2022). URL: https://arxiv.org/abs/2212.00968.

[2] E. Artemova, M. Zmeev, N. Loukachevitch, I. Rozhkov, T. Batura, V. Ivanov, E. Tutubalina, Runne-2022 shared task: Recognizing nested named entities, Komp'juternaja Lingvistika i Intellektual'nye Tehnologii 2022 (2022) 33 – 41. doi:10.28995/2075-7182-2022-21-33-41.

[3] N. Loukachevitch, E. Artemova, T. Batura, P. Braslavski, V. Ivanov, S. Manandhar, A. Pugachev, I. Rozhkov, A. Shelmanov, E. Tutubalina, et al., Nerel: a russian information extraction dataset with rich annotation for nested entities, relations, and wikidata entity links, Language Resources and Evaluation (2023) 1–37.

[4] N. 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, Bioinformatics (2023). URL: https://doi.org/10.1093/bioinformatics/btad161. doi:10.1093/bioinformatics/btad161, btad161.

[5] 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, Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 4626–4643. URL: https://aclanthology.org/2024.findings-naacl.288/. doi:10.18653/v1/2024.findings-naacl.288.

[6] 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, Association for Computational Linguistics, Online, 2021, pp. 4228–4238. URL: https://aclanthology.org/2021.naacl-main.334. doi:10.18653/v1/2021.naacl-main.334.

[7] A. Alekseev, Z. Miftahutdinov, E. Tutubalina, A. Shelmanov, V. Ivanov, V. Kokh, A. Nesterov, M. Avetisian, A. Chertok, S. Nikolenko, Medical crossing: a cross-lingual evaluation of clinical entity linking, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 4212–4220. URL: https://aclanthology.org/2022.lrec-1.447/.

[8] A. Nesterov, G. Zubkova, Z. Miftahutdinov, V. Kokh, E. Tutubalina, A. Shelmanov, A. Alekseev, M. Avetisian, A. Chertok, S. Nikolenko, Ruccon: clinical concept normalization in russian, in: Findings of the Association for Computational Linguistics: ACL 2022, 2022, pp. 239–245.

[9] A. Sakhovskiy, N. Semenova, A. Kadurin, E. Tutubalina, Graph-enriched biomedical entity representation transformer, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Springer Nature Switzerland, Cham, 2023, pp. 109–120.

[10] N. Loukachevitch, A. Sakhovskiy, E. Tutubalina, Biomedical concept normalization over nested entities with partial UMLS terminology in Russian, in: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 2383–2389. URL: https://aclanthology.org/2024.lrec-main.213/.

[11] V. Davydova, N. Loukachevitch, E. Tutubalina, Overview of BioNNE Task on Biomedical Nested Named Entity Recognition at BioASQ 2024, in: CLEF Working Notes, 2024.

[12] A. Nentidis, G. Katsimpras, A. Krithara, S. Lima-López, E. Farré-Maduell, M. Krallinger, N. Loukachevitch, V. Davydova, E. Tutubalina, G. Paliouras, Overview of BioASQ 2024: The

twelfth BioASQ challenge on Large-Scale Biomedical Semantic Indexing and Question Answering, in: L. Goeuriot, P. Mulhem, G. Quénot, D. Schwab, L. Soulier, G. Maria Di Nunzio, P. Galuščáková, A. García Seco de Herrera, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024), 2024.

[13] O. Bodenreider, The unified medical language system (UMLS): integrating biomedical terminology, Nucleic Acids Res. 32 (2004) 267–270. URL: https://doi.org/10.1093/nar/gkh061. doi:10.1093/NAR/GKH061.

[14] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, J. Tsujii, brat: a web-based tool for NLP-assisted text annotation, in: Proceedings of the Demonstrations Session at EACL 2012, Association for Computational Linguistics, Avignon, France, 2012.

[15] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference 1 (2018) 4171–4186. URL: http://arxiv.org/abs/1810.04805. arXiv:1810.04805.

[16] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, H. Poon, Domain-specific language model pretraining for biomedical natural language processing, ACM Transactions on Computing for Healthcare (HEALTH) 3 (2021) 1–23.

[17] D. Peña Gnecco, J. Serrano, E. Puertas, J. C. Martínez-Santos, Hybrid Re-ranking for Biomedical Entity Linking using SapBERT Embeddings: A High-Performance System for BioNNE-L 2025-1, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), CLEF 2025 Working Notes, 2025.

[18] Y. Liu, LYX_DMIIP_FDU at BioASQ 2025: Utilizing BERT embeddings for biomedical text mining, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), CLEF 2025 Working Notes, 2025.

[19] A. Burlova, Navigating Partial UMLS Terminology: GAT Embeddings and Confidence Analysis for Multilingual Concept Linking, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), CLEF 2025 Working Notes, 2025.

[20] C. Li, X. Zheng, S. Liu, BIBERT on Biomedical Nested Named Entity Linking at BioASQ 2025, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), CLEF 2025 Working Notes, 2025.

[21] F. Liu, I. Vulić, A. Korhonen, N. Collier, Learning domain-specialised representations for cross-lingual biomedical entity linking, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Association for Computational Linguistics, Online, 2021, pp. 565–574. URL: https://aclanthology.org/2021.acl-short.72/. doi:10.18653/v1/2021.acl-short.72.

[22] 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, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.

[23] A. van den Oord, Y. Li, O. Vinyals, Representation learning with contrastive predictive coding, ArXiv abs/1807.03748 (2018). URL: https://api.semanticscholar.org/CorpusID:49670925.

[24] A. D. Lain, C. Lee, S. E. Doneva, M. J. Rodríguez-Cubillos, E. Castagnari, T. I. Simpson, , J. M. Posma, Multilingual and Nested Biomedical Named Entity Normalisation via Candidate Retrieval and Lightweight Large Language Model Disambiguation, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), CLEF 2025 Working Notes, 2025.

[25] M. Sung, H. Jeon, J. Lee, J. Kang, Biomedical entity representations with synonym marginalization, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 3641–3650. URL: https://aclanthology.org/2020.acl-main.335/. doi:10.18653/v1/2020.acl-main.335.

[26] 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. arXiv:1810.04805.

[27] I. Beltagy, K. Lo, A. Cohan, Scibert: Pretrained language model for scientific text, in: EMNLP, 2019. arXiv:arXiv:1903.10676.

[28] S. Tedeschi, V. Maiorca, N. Campolungo, F. Cecconi, R. Navigli, WikiNEuRal: Combined neural and knowledge-based silver data creation for multilingual NER, in: Findings of the Association for Computational Linguistics: EMNLP 2021, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 2521–2533. URL: https://aclanthology.org/2021.findings-emnlp.215/. doi:10.18653/v1/2021.findings-emnlp.215.

[29] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, W. Wang, Language-agnostic BERT sentence embedding, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 878–891. URL: https://aclanthology.org/2022.acl-long.62/. doi:10.18653/v1/2022.acl-long.62.