BioASQ at CLEF2026: The fourteenth edition of the large-scale biomedical semantic indexing and question answering challenge

Anastasios Nentidis[1], Georgios Katsimpras[1], Anastasia Krithara[1], Martin Krallinger[2], Miguel Rodriguez Ortega[2], Eduard Rodriguez Lopez[2], Natalia Loukachevitch[3], Igor Rozhkov[3], Elena Tutubalina[4,5,6], Grigorios Tsoumakas[7], George Giannakoulas[7], Dimitris Dimitriadis[7], Alexandra Bekiaridou[8], Athanasios Samaras[7], Vasiliki Patsiou[7], Giorgio Maria Di Nunzio[9], Nicola Ferro[9], Stefano Marchesin[9], Marco Martinelli[9], Gianmaria Silvello[9], and Georgios Paliouras[1]

[1] National Center for Scientific Research "Demokritos", Athens, Greece
{tasosnent, gkatsibras, akrithara, paliourg}@iit.demokritos.gr
[2] Barcelona Supercomputing Center, Barcelona, Spain
{martin.krallinger, miguel.rodriguez, eduard.rodriguez}@bsc.es
[3] Moscow State University, Russia
louk_nat@mail.ru
[4] Sber AI, Russia
[5] Artificial Intelligence Research Institute, Russia
[6] Kazan Federal University, Russia
tutubalinaev@gmail.com
[7] Aristotle University of Thessaloniki, Greece
{greg,dndimitri}@csd.auth.gr, {g.giannakoulas, ath.samaras.as}@gmail.com
[8] Northwell Health, USA
ampekiaridou@gmail.com
[9] University of Padua, Italy
{giorgiomaria.dinunzio, nicola.ferro, stefano.marchesin, marco.martinelli, gianmaria.silvello}@unipd.it

Abstract. Over the past thirteen years, the large-scale biomedical semantic indexing and question-answering challenge (BioASQ) has consistently driven the advancement of methods and tools that enhance access to the rapidly growing body of scientific resources in the biomedical domain. BioASQ offers a unique common testbed where research teams worldwide can evaluate and compare innovative approaches for accessing biomedical knowledge by organizing shared tasks on an annual basis and providing respective benchmark datasets that represent the real information needs of biomedical experts.

The fourteenth version of BioASQ (BioASQ14) will be held as an evaluation Lab in the context of CLEF2026, providing six tasks: (i) *Task b* on biomedical semantic question answering. (ii) *Task Synergy* on question answering for developing biomedical topics. (iii) *Task MultiClinSum-2* on multilingual clinical summarization. (iv) *Task BioNNE-R* on nested relation extraction in Russian and English. (v) *Task ELCardioCC* on

clinical coding in cardiology. (vi) *Task GutBrainIE* on gut-brain interplay information extraction. Through its six shared tasks, BioASQ14 challenges the research community to develop methods that go beyond the current state of the art, fostering innovative approaches that enable efficient and precise access to biomedical knowledge while pushing the research frontier forward.

Keywords: Biomedical · Semantic Indexing · Question Answering

1 Introduction

BioASQ[1] [37] was introduced in 2012 as a series of international challenges and workshops on biomedical semantic indexing and question answering (QA). Since then, BioASQ has offered a range of shared tasks on biomedical information access, machine learning, information retrieval, and information extraction, and has welcomed participation from more than 150 teams from 36 distinct countries. The benchmark datasets and corresponding open-source infrastructure developed for BioASQ tasks enable researchers in biomedical information access systems to work on the same realistic scenarios and compare the performance of their approaches. In addition, the respective annual BioASQ workshops allow the participants to present and discuss their methods with the research community [8, 11, 30–32, 19]. By consistently recognizing the most successful approaches in each shared task and sub-task, BioASQ drives the development of systems that surpass previous methods. These advances in biomedical semantic indexing and QA are progressively enabling more accurate retrieval of biomedical knowledge and contributing to the enhancement of healthcare services.

2 BioASQ evaluation lab 2026

The fourteenth BioASQ challenge (BioASQ14), to be held as part of the Conference and Labs of the Evaluation Forum (CLEF) 2026[2], will be structured into six tasks that are central to accessing biomedical knowledge and answering biomedical questions: (i) *Task b*[3] on the processing of biomedical questions, the generation of answers, and the retrieval of supporting material, (ii) *Task Synergy* on biomedical QA for developing topics, such as COVID-19, where resources are not necessarily sufficient to provide definite answers, (iii) *Task MultiClinSum-2* on Multilingual Clinical Summarization. (iv) *Task BioNNE-R*: on the automated relation extraction between nested named entities in Russian and English. (v) *Task ELCardioCC* on clinical coding of Greek cardiology discharge letters. (vi) *Task GutBrainIE* on extracting and linking knowledge from the scientific literature on the gut-brain interplay. As *Task b* and *Task Synergy* have also been

[1] http://www.bioasq.org

[2] https://clef2026.clef-initiative.eu/

[3] Since the first BioASQ, the task on biomedical semantic indexing has been called *Task a* and the task on QA *Task b*, for brevity. *Task a* was completed in 2020 [9].

organized in the context of previous editions of the BioASQ challenge [22, 21, 27, 25], we refer to their current version, in the context of BioASQ14, as *task 14b*, *task Synergy 14* respectively. Detailed guidelines for participating in each BioASQ task are available online[4].

2.1 Task 14b: Biomedical question answering

BioASQ *task 14b* is conducted in three phases. In Phase A, systems are given biomedical questions in English and must retrieve relevant material that is PubMed documents and snippets. In Phase A+, systems are required to provide both 'ideal' and 'exact' answers for these questions. The 'ideal' answer is a paragraph-length summary, regardless of the question type but the form of the 'exact' answer varies by question type: it can be a *yes* or *no* (yes/no type), an entity name such as a gene or drug (factoid type), or a list of entity names (list type)[5]. Finally, in Phase B, systems are supplied with relevant material selected by the BioASQ experts and must provide updated 'ideal' and 'exact' answers based on this additional information. The results from the previous version of the task [25] reveal that state-of-the-art LLM-based systems can provide correct answers for yes/no questions, even without access to ground truth relevant material (Top maF1 0.97 ±0.03 in Phase A+). For factoid and list questions, on the other hand, there is more room for improvement (top factoid MRR 0.64 ±0.11, top list F1 0.57 ±0.02 in Phase B), especially without access to the ground truth relevant material (top factoid MRR 0.64 ±0.11 and top list F1 0.57 ±0.02 in Phase A+). Therefore, the retrieval of relevant material appears to be a significant persisting challenge (top document MAP 0.58 ±0.08 and top snippet f1 0.24 ±0.02 in Phase A)[6].

About 300 new biomedical questions annotated with golden documents, snippets, and answers ('exact' and 'ideal') will be developed for testing. In addition, a training set of about 5,700 biomedical questions, accompanied by answers and supporting evidence (documents and snippets), will be available from previous versions of the tasks, as a unique resource for the development of question-answering systems [20, 10]. The evaluation in *task 14b* is done manually by the experts who assess each system response and automatically by employing a variety of established evaluation measures [16] as in *task 13b* [25].

2.2 Task Synergy 14: Question answering for developing topics

In 2020, we introduced the BioASQ *task Synergy* to advance research on emerging biomedical topics, such as COVID-19 [11, 12]. Unlike the original *task b*, *task Synergy* is structured as a continuous dialogue, enabling experts to pose open questions on developing topics for which a definitive answer may not be

[4] https://participants-area.bioasq.org/

[5] For questions of type summary, no 'exact' answer is defined, only an 'ideal' answer.

[6] The reported performance of top participating systems has been averaged across the four test batches of BioASQ13

known in advance. Systems provide relevant material, including documents and snippets, along with answers, which are then assessed by the BioASQ experts. Feedback from these assessments is provided back to the systems after each round to refine their responses to the same questions. This iterative process continues across multiple rounds, incorporating new material based on updates to the original document resource [26]. Since 2023, this evolving document resource is PubMed [21]. During task Synergy 13 [25], the systems managed to identify enough relevant material to provide an answer to about 80% of the questions and provide at least one ideal answer of ground-truth quality for about 47% of them.

A training dataset for *task Synergy* is already available from previous versions of the task [25, 24, 23, 29, 28]. It consists of approximately 400 questions on developing topics, with incremental annotations including relevant material and answers. In addition, this set will be extended with about fifty new open questions on developing health topics and incremental feedback for the systems after each of the four rounds of *task Synergy 14*. Meanwhile, any existing questions that remain relevant may be enriched with more recent material and more up-to-date answers as well. The primary goal of *task Synergy 14* is to support experts in incrementally advancing the understanding of evolving health topics and in discovering new solutions. However, the same evaluation measures used in *task 14b* will also apply for completeness. For the information retrieval part, only new material will be considered per round (*residual collection evaluation* [36]).

2.3 Task MultiClinSum-2: Multilingual Clinical Summarization

There is an increasing demand for automated tools that can efficiently process and summarise clinical narratives written in multiple languages. Clinical case reports and medical records are accumulating rapidly and often contain lengthy, jargon-heavy descriptions that are time-consuming for clinicians and researchers to understand and extract relevant information. While Large Language Models (LLMs) have made significant progress in clinical summarisation, their performance across languages, data domains and summarisation styles remains insufficiently benchmarked.

The MultiClinSum-2 task extends the previous edition by expanding both the dataset and evaluation scope for multilingual clinical summarization. As in the first edition, participants will generate summaries for full-length clinical case reports written in English, Spanish, French, and Portuguese. In addition, new languages such as Italian, Swedish, Czech, or Dutch are being considered for addition, further broadening the linguistic coverage of the task. The summaries, which will be released as Gold Standard, are based on the original author-provided summaries of the corresponding full clinical case report descriptions. These summaries are not supposed to be readable by laypersons, as this is not the main focus of the task, but rather by medical-related experts to cover key clinical insight, such as patient characteristics (age, gender, etc.), reason for consultation, anamnesis and physical exploration, diagnosis, treatment, and outcome. Beyond traditional lexical and semantic metrics such as ROUGE-2 and

BERTScore, this second edition will introduce factual consistency (FC) assessment and LLM-as-a-Judge evaluation, where large language models may assist in ranking or validating summary quality. As clinical case reports share common structure with medical discharge summaries, MultiClinSum-2 aims to provide insights not only for publication-style narratives but also for real-world healthcare applications.

2.4 Task BioNNE-R: Nested Relation Extraction in Russian and English

The BioNNE-R shared task addresses the NLP challenge of relation extraction involving nested named entities, i.e. entities that contain other entities within their boundaries. Relations between nested entities may cross entity boundaries, connecting shorter entities within longer ones, making detection more challenging. For example, *blood system tumor* includes internal entities *blood*, *blood system*, and *tumor*. Several internal relations within *blood system tumor* can be identified: *blood* is a part of *blood system*; *blood system tumor* is a subclass of *tumor*; *blood system tumor* affects *blood system*. Moreover, internal entities can also relate to external entities as *blood system* is a part of *patient*, mentioned separately. Capturing such nested relations enables more comprehensive information extraction, yet the annotation and extraction of nested relations remain insufficiently studied.

Following earlier shared tasks on nested entity processing *BioNNE 2024* [4] and *BioNNE-L 2025* [35] tasks, the evaluation framework is divided as follows: 1. Track Language-oriented: Participants in this track must develop a model for nested relation extraction in a target language (English or Russian). 2. Track Bilingual: Participants must train a single model using training data for both Russian and English languages. The task data will be based on the NEREL-BIO dataset [15], which includes annotated mentions of disorders, anatomical structures, chemicals, diagnostic procedures, and other biomedical entities.

2.5 Task ELCardioCC: Clinical Coding in Cardiology

Clinical coding is the process of translating clinical documentation from patient health records into standardized coding systems [5]. This translation facilitates the generation of reliable datasets for epidemiological research, health service planning, and statistical evaluation. The procedure is inherently complex, requiring significant cognitive effort and meticulous adherence to coding standards to maintain uniformity and validity across records. Nevertheless, the majority of existing research has concentrated on clinical text written in English, resulting in a notable underrepresentation of other languages, including Greek. This linguistic imbalance limits the generalizability of current findings and creates challenges for developing language-specific resources, such as annotated corpora, terminologies, and natural language processing tools. Addressing this gap is essential for enabling equitable access to advanced clinical text mining techniques across diverse healthcare systems.

The *ELCardioCC* 2026 shared task concerns the automatic assignment of cardiology-related ICD-10 codes to hospital discharge letters at the document level. The task can be approached in two ways: either by combining Named Entity Recognition (NER) and Entity Linking (EL) techniques, or by formulating it as a multi-label classification (MLC) problem. To support both directions, we introduce a mixed dataset consisting of 1,500 documents annotated at both the document and mention level, suitable for NER+EL approaches, and 1,000 documents annotated only at the document level, intended for MLC approaches. In total, the dataset comprises 2,500 documents for training and development and 500 documents for testing. Finally, the coding standards will be available to participants.

2.6 Task GutBrainIE: Gut-Brain Interplay Information Extraction

Recent evidence suggests a connection between neurological and gut disorders that may play a critical role in mental health-related disorders or diseases like Multiple Sclerosis, Parkinson's, and Alzheimer's [1–3, 7]. The *GutBrainIE* task aims to foster the development of Information Extraction (IE) systems that support experts by automatically extracting and linking knowledge from scientific literature, facilitating the understanding of gut-brain interplay and its role in neurological diseases. The task is divided into two main subtasks. In the first, participants are provided with PubMed abstracts discussing the gut-brain interplay and asked to identify named entities and link them to the corresponding concepts in reference biomedical resources (e.g., UMLS). In the second subtask, participants are required to identify relations between pairs of extracted entities, assigning the appropriate relation predicate that connects them. For both subtasks, submitted runs are evaluated using Precision, Recall, and F1 measures against gold annotations created by domain experts. Last year's results highlight a clear difficulty gap between the two subtasks: half of the participating teams achieved strong effectiveness in named entity recognition (F1 ≥ 0.8), while relation extraction saw the best systems scoring in the 0.34–0.46 F1 range. Top-performing approaches relied on supervised fine-tuning of classification heads on top of transformer models pre-trained on biomedical corpora, while LLM-based approaches showed limited effectiveness [17].

The GutBrainIE dataset comprises over 1,600 PubMed abstracts annotated with entity mentions, concept links, and relations. We focus on abstracts only as they summarize the most important information in each paper and typically give a higher density of entity mentions of interest. The entity schema covers general biomedical categories (e.g., *bacteria*, *drug*), gut-brain specific ones (e.g., *microbiome*), and experimental concepts (e.g., *statistical technique*). About relations, we defined a total of 17 predicates, resulting in 55 possible relation triples *(head entity, predicate, tail entity)*. The dataset is split into training, validation, and test sets. Together with the dataset, we provide reference resources used for entity linking and annotation guidelines.

2.7 BioASQ datasets and tools

BioASQ offers an ecosystem of publicly available datasets[7] and tools[8], including a range of evaluation measures [16] and the BioASQ Annotation Tool [33] for the development of QA datasets. In addition to the unique datasets provided for this year's six tasks, BioASQ also offers: i) a benchmark dataset of more than 16.2 million articles on biomedical semantic indexing (*task a*) [9], ii) the *task MESINESP* datasets [6, 34] of more than 300K articles, on medical semantic indexing in Spanish, iii) the *task DisTEMIST* [18], *task MedProcNER* [13], and *task MultiCardioNER* [14] on medical information extraction from clinical case documents, iv) the datasets of the tasks *task BioNNE* [4] and *task BioNNE-L* [35] on nested named entity recognition and linking.

3 Conclusions

BioASQ promotes the exchange and integration of ideas by offering shared tasks on semantic indexing and question answering, and providing unique, realistic datasets and evaluation tools for research on biomedical information access methods. This, in turn, accelerates progress in the field, as reflected in the steady improvement of task scores [20, 22, 21, 27]. A notable example is BioASQ's role in supporting the adoption of fully automated MeSH indexing at NLM [9]. Likewise, we anticipate that this edition of BioASQ will enable participating teams to further advance the six open tasks offered this year.

Acknowledgments This edition of BioASQ was sponsored by Ovid Technologies, Inc. This research was funded by the Spanish National BARITONE project (TED2021-129974B-C22). This work is also supported by the European Union's Horizon Europe Co-ordination & Support Action under Grant Agreement No 101080430 (AI4HF), as well as Grant Agreement No 101057849 (DataTool4Heartproject). The work on the *BioNNE-R* task was supported by the Russian Science Foundation [grant number 23-11-00358]. This work is partially supported by the HEREDITARY Project, as part of the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Appleton, J.: The gut-brain axis: influence of microbiota on mood and mental health. Integrative Medicine: A Clinician's Journal **17**(4), 28 (2018)

[7] http://participants-area.bioasq.org/datasets
[8] https://github.com/bioasq

2. Carabotti, M., Scirocco, A., Maselli, M.A., Severi, C.: The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. Annals of gastroenterology: quarterly publication of the Hellenic Society of Gastroenterology **28**(2), 203 (2015)
3. Cryan, J.F., O'Riordan, K.J., Sandhu, K., Peterson, V., Dinan, T.G.: The gut microbiome in neurological disorders. The Lancet Neurology **19**(2), 179–194 (2020)
4. Davydova, V., Loukachevitch, N., Tutubalina, E.: Overview of bionne task on biomedical nested named entity recognition at bioasq 2024. In: CLEF Working Notes (2024), https://ceur-ws.org/Vol-3740/paper-03.pdf
5. Dong, H., Falis, M., Whiteley, W., Alex, B., Matterson, J., Ji, S., Chen, J., Wu, H.: Automated clinical coding: what, why, and where we are? NPJ digital medicine **5**(1), 159 (2022). https://doi.org/10.1038/s41746-022-00705-7
6. Gasco, L., Nentidis, A., Krithara, A., Estrada-Zavala, D., Murasaki, R.T., Primo-Peña, E., Bojo Canales, C., Paliouras, G., Krallinger, M., et al.: Overview of BioASQ 2021-MESINESP track. Evaluation of advance hierarchical classification techniques for scientific literature, patents and clinical trials. CEUR Workshop Proceedings (2021), https://ceur-ws.org/Vol-2936/paper-11.pdf
7. Ghaisas, S., Maher, J., Kanthasamy, A.: Gut microbiome in health and disease: Linking the microbiome–gut–brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases. Pharmacology & therapeutics **158**, 52–62 (2016). https://doi.org/10.1016/j.pharmthera.2015.11.012
8. Krallinger, M., Krithara, A., Nentidis, A., Paliouras, G., Villegas, M.: BioASQ at CLEF2020: Large-scale biomedical semantic indexing and question answering. In: European Conference on Information Retrieval. pp. 550–556. Springer (2020). https://doi.org/10.1007/978-3-030-45442-5_71
9. Krithara, A., Mork, J.G., Nentidis, A., Paliouras, G.: The road from manual to automatic semantic indexing of biomedical literature: a 10 years journey. Frontiers in Research Metrics and Analytics **8** (2023). https://doi.org/10.3389/frma.2023.1250930
10. Krithara, A., Nentidis, A., Bougiatiotis, K., Paliouras, G.: BioASQ-QA: A manually curated corpus for Biomedical Question Answering. Scientific Data **10**(1), 170 (2023). https://doi.org/10.1038/s41597-023-02068-4
11. Krithara, A., Nentidis, A., Paliouras, G., Krallinger, M., Miranda, A.: BioASQ at CLEF2021: Large-Scale Biomedical Semantic Indexing and Question Answering. In: European Conference on Information Retrieval. pp. 624–630. Springer (2021). https://doi.org/10.1007/978-3-030-72240-1_73
12. Krithara, A., Nentidis, A., Vandorou, E., Katsimpras, G., Almirantis, Y., Arnal, M., Bunevicius, A., Farre-Maduell, E., Kassiss, M., Konstantakos, V., Matis-Mitchell, S., Polychronopoulos, D., Rodriguez-Pascual, J., Samaras, E.G., Samiotaki, M., Sanoudou, D., Vozi, A., Paliouras, G.: BioASQ Synergy: a dialogue between question-answering systems and biomedical experts for promoting COVID-19 research. Journal of the American Medical Informatics Association p. ocae232 (08 2024). https://doi.org/10.1093/jamia/ocae232
13. Lima-López, S., Farré-Maduell, E., Gascó, L., , Nentidis, A., Krithara, A., Katsimpras, G., Paliouras, G., Krallinger, M.: Overview of MedProcNER Task on Medical Procedure Detection and Entity Linking at BioASQ 2023. In: CEUR Workshop Proceedings (2023), https://ceur-ws.org/Vol-3497/paper-002.pdf
14. Lima-López, S., Farré-Maduell, E., Rodríguez-Miret, J., Rodríguez-Ortega, M., Lilli, L., Lenkowicz, J., Ceroni, G., Kossoff, J., Shah, A., Nentidis, A., Krithara, A., Katsimpras, G., Paliouras, G., Krallinger, M.: Overview of multicardioner task

at bioasq 2024 on medical specialty and language adaptation of clinical ner systems for spanish, english and italian. In: Conference and Labs of the Evaluation Forum (2024), https://ceur-ws.org/Vol-3740/paper-02.pdf
15. Loukachevitch, N., Manandhar, S., Baral, E., Rozhkov, I., Braslavski, P., Ivanov, V., Batura, T., Tutubalina, E.: Nerel-bio: a dataset of biomedical abstracts annotated with nested named entities. Bioinformatics **39**(4), btad161 (2023). https://doi.org/10.1093/bioinformatics/btad161
16. Malakasiotis, P., Pavlopoulos, I., Androutsopoulos, I., Nentidis, A.: Evaluation measures for task b. Tech. rep., Tech. rep. BioASQ (2018), http://participants-area.bioasq.org/Tasks/b/eval_meas_2018
17. Martinelli, M., Silvello, G., Bonato, V., Di Nunzio, G., Ferro, N., Irrera, O., Marchesin, S., Menotti, L., Vezzani, F., et al.: Overview of gutbrainie@ clef 2025: Gutbrain interplay information extraction. In: Faggioli, G., Ferro, N., Rosso, P., Spina, D. (eds.) CLEF 2025 Working Notes. vol. 4038, pp. 65–98 (2025), https://ceur-ws.org/Vol-4038/paper_5.pdf
18. Miranda-Escalada, A., Gascó, L., Lima-López, S., Farré-Maduell, E., Estrada, D., Nentidis, A., Krithara, A., Katsimpras, G., Paliouras, G., Krallinger, M.: Overview of DisTEMIST at BioASQ: Automatic detection and normalization of diseases from clinical texts: results, methods, evaluation and multilingual resources. In: Working Notes of Conference and Labs of the Evaluation (CLEF) Forum. CEUR Workshop Proceedings (2022), https://ceur-ws.org/Vol-3180/paper-11.pdf
19. Nentidis, A., Katsimpras, G., Krithara, A., Krallinger, M., Ortega, M.R., Loukachevitch, N., Sakhovskiy, A., Tutubalina, E., Tsoumakas, G., Giannakoulas, G., Bekiaridou, A., Samaras, A., Di Nunzio, G.M., Ferro, N., Marchesin, S., Menotti, L., Silvello, G., Paliouras, G.: BioASQ at CLEF2025: The Thirteenth Edition of the Large-Scale Biomedical Semantic Indexing and Question Answering Challenge. In: Advances in Information Retrieval. pp. 407–415. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-88720-8_61
20. Nentidis, A., Katsimpras, G., Krithara, A., Krallinger, M., Rodríguez-Ortega, M., Rodriguez-López, E., Loukachevitch, N., Sakhovskiy, A., Tutubalina, E., Dimitriadis, D., Tsoumakas, G., Giannakoulas, G., Bekiaridou, A., Samaras, A., Nunzio, G.M.D., Ferro, N., Marchesin, S., Martinelli, M., Silvello, G., Paliouras, G.: Overview of BioASQ 2025: The Thirteenth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: Carrillo-de Albornoz, J., García Seco de Herrera, A., Gonzalo, J., Plaza, L., Mothe, J., Piroi, F., Rosso, P., Spina, D., Faggioli, G., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. pp. 173–198. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-031-88720-8_61
21. Nentidis, A., Katsimpras, G., Krithara, A., Lima López, S., Farré-Maduell, E., Gasco, L., Krallinger, M., Paliouras, G.: Overview of BioASQ 2023: The Eleventh BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Giachanou, A., Li, D., Aliannejadi, M., Vlachos, M., Faggioli, G., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. pp. 227–250. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-42448-9_19
22. Nentidis, A., Katsimpras, G., Krithara, A., Lima-López, S., Farré-Maduell, E., Krallinger, M., Loukachevitch, N., Davydova, V., Tutubalina, E., Paliouras, G.: Overview of BioASQ 2024: The twelfth BioASQ challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: Goeuriot, L., Mulhem, P., Quénot, G., Schwab, D., Soulier, L., Maria Di Nunzio, G., Galuščáková, P.,

García Seco de Herrera, A., Faggioli, G., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024) (2024). https://doi.org/10.1007/978-3-031-71908-0_1

23. Nentidis, A., Katsimpras, G., Krithara, A., Paliouras, G.: Overview of BioASQ Tasks 11b and Synergy11 in CLEF2023. Working Notes of CLEF (2023), https://ceur-ws.org/Vol-3497/paper-003.pdf
24. Nentidis, A., Katsimpras, G., Krithara, A., Paliouras, G.: Overview of BioASQ Tasks 12b and Synergy12 in CLEF2024. In: Faggioli, G., Ferro, N., Galuščáková, P., García Seco de Herrera, A. (eds.) Working Notes of CLEF (2024), https://ceur-ws.org/Vol-3740/paper-01.pdf
25. Nentidis, A., Katsimpras, G., Krithara, A., Paliouras, G.: Overview of BioASQ Tasks 13b and Synergy13 in CLEF2025. In: Faggioli, G., Ferro, N., Rosso, P., Spina, D. (eds.) CLEF 2025 Working Notes (2025), https://ceur-ws.org/Vol-4038/paper_1.pdf
26. Nentidis, A., Katsimpras, G., Vandorou, E., Krithara, A., Gascó, L., Krallinger, M., Paliouras, G.: Overview of BioASQ 2021: The Ninth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: 12th International Conference of the CLEF Association. Lecture Notes in Computer Science, vol. 12880, pp. 239–263 (2021). https://doi.org/10.1007/978-3-030-85251-1_18
27. Nentidis, A., Katsimpras, G., Vandorou, E., Krithara, A., Miranda-Escalada, A., Gasco, L., Krallinger, M., Paliouras, G.: Overview of BioASQ 2022: The Tenth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13390 LNCS, pp. 337–361 (oct 2022). https://doi.org/10.1007/978-3-031-13643-6_22
28. Nentidis, A., Katsimpras, G., Vandorou, E., Krithara, A., Paliouras, G.: Overview of BioASQ Tasks 9a, 9b and Synergy in CLEF2021. In: Proceedings of the 9th BioASQ Workshop A challenge on large-scale biomedical semantic indexing and question answering. CEUR Workshop Proceedings (2021), http://ceur-ws.org/Vol-2936/paper-10.pdf
29. Nentidis, A., Katsimpras, G., Vandorou, E., Krithara, A., Paliouras, G.: Overview of BioASQ Tasks 10a, 10b and Synergy10 in CLEF2022. In: CEUR Workshop Proceedings. vol. 3180, pp. 171–178 (2022), https://ceur-ws.org/Vol-3180/paper-10.pdf
30. Nentidis, A., Krithara, A., Paliouras, G., Farre-Maduell, E., Lima-Lopez, S., Krallinger, M.: BioASQ at CLEF2023: The Eleventh Edition of the Large-Scale Biomedical Semantic Indexing and Question Answering Challenge. In: Advances in Information Retrieval: ECIR 2023, Dublin, Ireland, April 2–6, 2023, Proceedings, Part III. pp. 577–584. Springer (2023). https://doi.org/10.1007/978-3-031-28241-6_66
31. Nentidis, A., Krithara, A., Paliouras, G., Gasco, L., Krallinger, M.: BioASQ at CLEF2022: The Tenth Edition of the Large-scale Biomedical Semantic Indexing and Question Answering Challenge. In: Advances in Information Retrieval: 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10–14, 2022, Proceedings, Part II. Springer, Springer (2022). https://doi.org/10.1007/978-3-030-99739-7_53
32. Nentidis, A., Krithara, A., Paliouras, G., Krallinger, M., Sanchez, L.G., Lima, S., Farre, E., Loukachevitch, N., Davydova, V., Tutubalina, E.: BioASQ at CLEF2024: The Twelfth Edition of the Large-Scale Biomedical Semantic In-

dexing and Question Answering Challenge. In: Advances in Information Retrieval. Springer Nature Switzerland, Springer Nature Switzerland, Cham (2024), https://link.springer.com/chapter/10.1007/978-3-031-56069-9_67
33. Ngomo, A.C.N., Heino, N., Speck, R., Ermilov, T., Tsatsaronis, G.: Annotation tool. Project deliverable D3.3 (02/2013 2013), http://www.bioasq.org/sites/default/files/PublicDocuments/2013-D3.3-AnnotationTool.pdf
34. Rodriguez-Penagos, C., Nentidis, A., Gonzalez-Agirre, A., Asensio, A., Armengol-Estapé, J., Krithara, A., Villegas, M., Paliouras, G., Krallinger, M.: Overview of MESINESP8, a Spanish Medical Semantic Indexing Task within BioASQ 2020 (2020), https://ceur-ws.org/Vol-2696/paper_269.pdf
35. Sakhovskiy, A., Loukachevitch, N., Tutubalina, E.: Overview of the BioASQ BioNNE-L Task on Biomedical Nested Entity Linking in CLEF 2025. In: Faggioli, G., Ferro, N., Rosso, P., Spina, D. (eds.) CLEF 2025 Working Notes (2025), https://ceur-ws.org/Vol-4038/paper_3.pdf
36. Salton, G., Buckley, C.: Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science **41**(4), 288–297 (jun 1990). https://doi.org/10.1002/(SICI)1097-4571(199006)41:4¡288::AID-ASI8¿3.0.CO;2-H
37. Tsatsaronis, G., Balikas, G., Malakasiotis, P., Partalas, I., Zschunke, M., Alvers, M.R., Weissenborn, D., Krithara, A., Petridis, S., Polychronopoulos, D., Almirantis, Y., Pavlopoulos, J., Baskiotis, N., Gallinari, P., Artieres, T., Ngonga, A., Heino, N., Gaussier, E., Barrio-Alvers, L., Schroeder, M., Androutsopoulos, I., Paliouras, G.: An overview of the BioASQ large-scale biomedical semantic indexing and question answering competition. BMC Bioinformatics **16**, 138 (2015). https://doi.org/10.1186/s12859-015-0564-6