

Resource Summary Report

Generated by SPARC SAWG on Sep 17, 2026

BioCreative

RRID:SCR_006311

Type: Tool

Proper Citation

BioCreative (RRID:SCR_006311)

Resource Information

URL: <http://biocreative.sourceforge.net/>

Proper Citation: BioCreative (RRID:SCR_006311)

Description: Community-wide effort (Challenge) for evaluating text mining and information extraction systems applied to the biological domain. It is focused on the comparison of methods and the community assessment of scientific progress, rather than on the purely competitive aspects. There is a considerable difficulty in constructing suitable gold standard data for training and testing new information extraction systems which handle life science literature. Thus the data sets derived from the BioCreAtIvE challenge - because they have been examined by biological database curators and domain experts - serve as useful resources for the development of new applications as well as helping to improve existing ones. Two main issues are addressed at BioCreAtIvE, both concerned with the extraction of biologically relevant and useful information from the literature. The first one is concerned with the detection of biologically significant entities (names) such as gene and protein names and their association to existing database entries. The second one is concerned with the detection of entity-fact associations (e.g. protein - functional term associations).

Abbreviations: BioCreAtIvE

Synonyms: BioCreAtIvE challenge evaluation, Critical Assessment of Information Extraction systems in Biology, BioCreAtIvE - Critical Assessment for Information Extraction in Biology

Resource Type: benchmark, challenge, data or information resource, data set, portal, topical portal

Keywords: text mining, information extraction, literature, crowdsourcing, bioinformatics, natural language processing, challenge, benchmark, evaluation, gold standard, FASEB list

Resource Name: BioCreative

Resource ID: SCR_006311

Alternate IDs: nlx_151992

Record Creation Time: 20220129T080235+0000

Record Last Update: 20260912T195638+0000

Ratings and Alerts

No rating or validation information has been found for BioCreative.

No alerts have been found for BioCreative.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 163 mentions in open access literature.

Listed below are recent publications. The full list is available at [SPARC SAWG](#) .

Alhassan A, et al. (2026) Recognition and linking of discontinuous named entities in healthcare: a comparative performance analysis. *Frontiers in digital health*, 8, 1758921.

Tsueng G, et al. (2026) The NIAID Discovery Portal: a unified search engine for infectious and immune-mediated disease datasets. *mSystems*, 11(2), e0127025.

Islamaj R, et al. (2026) Overview of the MedHopQA track at BioCreative IX: track description, participation and evaluation of systems for multi-hop medical question answering. *ArXiv*.

Happi Happi BG, et al. (2025) GRU-SCANET: unleashing the power of GRU-based sinusoidal capture network for precision-driven named entity recognition. *Bioinformatics advances*, 5(1), vbaf096.

Alhassan A, et al. (2025) Generative Models and Sentence Transformers for the Recognition and Normalization of Continuous and Discontinuous Phenotype Mentions: Model Development and Evaluation. *JMIR medical informatics*, 13, e68558.

Shang X, et al. (2025) Benchmarking large language models for genomic knowledge with GeneTuring. *Briefings in bioinformatics*, 26(5).

Baltoumas FA, et al. (2025) Darling (v2.0): Mining disease-related databases for the detection of biomedical entity associations. *Computational and structural biotechnology journal*, 27, 2626.

Caufield JH, et al. (2024) Structured Prompt Interrogation and Recursive Extraction of Semantics (SPIRES): a method for populating knowledge bases using zero-shot learning. *Bioinformatics* (Oxford, England), 40(3).

Ogrinc M, et al. (2024) Zero-shot evaluation of ChatGPT for food named-entity recognition and linking. *Frontiers in nutrition*, 11, 1429259.

Hala S, et al. (2024) The emergence of highly resistant and hypervirulent *Klebsiella pneumoniae* CC14 clone in a tertiary hospital over 8 years. *Genome medicine*, 16(1), 58.

Aldahdooh J, et al. (2024) Mining drug-target interactions from biomedical literature using chemical and gene descriptions-based ensemble transformer model. *Bioinformatics advances*, 4(1), vbae106.

Zheng D, et al. (2024) Biomedical named entity recognition based on multi-cross attention feature fusion. *PloS one*, 19(5), e0304329.

Liu S, et al. (2024) Evaluating Medical Entity Recognition in Health Care: Entity Model Quantitative Study. *JMIR medical informatics*, 12, e59782.

Vollmar M, et al. (2024) Dataset from a human-in-the-loop approach to identify functionally important protein residues from literature. *Scientific data*, 11(1), 1032.

Lai PT, et al. (2024) EnzChemRED, a rich enzyme chemistry relation extraction dataset. *Scientific data*, 11(1), 982.

Liu K, et al. (2024) Conductive and alignment-optimized porous fiber conduits with electrical stimulation for peripheral nerve regeneration. *Materials today. Bio*, 26, 101064.

Groza T, et al. (2024) FastHPOCR: pragmatic, fast, and accurate concept recognition using the human phenotype ontology. *Bioinformatics* (Oxford, England), 40(7).

Jonker RAA, et al. (2024) Multi-head CRF classifier for biomedical multi-class named entity recognition on Spanish clinical notes. *Database : the journal of biological databases and curation*, 2024.

Madan S, et al. (2024) Dataset of miRNA-disease relations extracted from textual data using transformer-based neural networks. *Database : the journal of biological databases and curation*, 2024.

Mehryary F, et al. (2024) STRING-ing together protein complexes: corpus and methods for extracting physical protein interactions from the biomedical literature. *Bioinformatics* (Oxford, England), 40(9).