

Resource Summary Report

Generated by SPARC SAWG on Aug 9, 2026

GENIA Project: Mining literature for knowledge in molecular biology

RRID:SCR_007990

Type: Tool

Proper Citation

GENIA Project: Mining literature for knowledge in molecular biology (RRID:SCR_007990)

Resource Information

URL: <http://www.nactem.ac.uk/genia/>

Proper Citation: GENIA Project: Mining literature for knowledge in molecular biology (RRID:SCR_007990)

Description: Resources and tools from a project to automatically extract useful information from texts written by scientists to help overcome the problems caused by information overload. The primary annotated resource created is the GENIA corpus, a collection of biomedical literature which consists of multiple layers of annotation, encompassing both syntactic and semantic annotation. The project also created or coordinated the annotation of multiple other corpus resources. Additionally, a rich set of automatic tools are available for various annotation tasks, most trained on various parts of the GENIA corpus annotations. The GENIA corpus was developed to provide a reference material for the development of bio-TM systems. The corpus currently contains 1,999 Medline abstracts which were collected using the three MeSH terms, human, blood cells, and transcription factors. The corpus has been annotated with various levels of linguistic and semantic information. The GENIA corpus includes the following: * POS annotation * Treebank * Coreference Annotation * Term annotation * Event annotation * Relation annotation * Cellular localization * Disease-Gene association * Pathway corpus The GENIA Project initiated the BioNLP Shared Task series and has organized a number of tasks in three different shared task events, many using resources based on GENIA Corpus annotations. Tools include: * XConc suite: a collection of XML-based tools which are integrated to support the corpus development and annotation.

Abbreviations: GENIA

Resource Type: software resource

Keywords: annotation, biomedical, computational linguistics, text mining, literature, molecular biology, syntactic annotation, semantic annotation, syntax, semantics, information extraction, blood cell, transcription factor, protein interaction, task

Funding: Japanese Ministry of Education Culture Sports Science and Technology MEXT ; Japan Science and Technology Agency

Resource Name: GENIA Project: Mining literature for knowledge in molecular biology

Resource ID: SCR_007990

Alternate IDs: nif-0000-06689

Old URLs: <http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/home/wiki.cgi?page=GENIA+Project>

Record Creation Time: 20220129T080244+0000

Record Last Update: 20260808T185911+0000

Ratings and Alerts

No rating or validation information has been found for GENIA Project: Mining literature for knowledge in molecular biology.

No alerts have been found for GENIA Project: Mining literature for knowledge in molecular biology.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 2 mentions in open access literature.

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

Ayd??n F, et al. (2017) Automatic query generation using word embeddings for retrieving passages describing experimental methods. Database : the journal of biological databases and curation, 2017.

Tudor CO, et al. (2012) The eFIP system for text mining of protein interaction networks of phosphorylated proteins. Database : the journal of biological databases and curation, 2012, bas044.