

Resource Summary Report

Generated by [SPARC SAWG](#) on Jul 28, 2026

GREC Corpus

RRID:SCR_006719

Type: Tool

Proper Citation

GREC Corpus (RRID:SCR_006719)

Resource Information

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

Proper Citation: GREC Corpus (RRID:SCR_006719)

Description: A semantically annotated corpus of 240 MEDLINE abstracts (167 on the subject of E. coli species and 73 on the subject of the Human species) intended for training information extraction (IE) systems and/or resources which are used to extract events from biomedical literature. The corpus has been manually annotated with events relating to gene regulation by biologists. Each event is centered on either a verb (e.g. transcribe) or nominalized verb (e.g. transcription) and annotation consists of identifying, as exhaustively as possible, the structurally-related arguments of the verb or nominalized verb within the same sentence. Each event argument is then assigned the following information: * A semantic role from a fixed set of 13 roles which are tailored to the biomedical domain. * A biomedical concept type (where appropriate). The corpus is available for download in 2 formats: * A standoff format, based on the BioNLP'09 Shared Task format * An XML format, based on the GENIA event annotation format

Abbreviations: GREC

Synonyms: Gene Event Regulation Corpus

Resource Type: training set

Defining Citation: [PMID:19852798](#)

Keywords: annotation, information extraction, text mining, semantic role, semantic search, gene, computational linguistics, gene regulation

Funding: JISC

Availability: Creative Commons Attribution-NonCommercial-ShareAlike License, v3 Unported, For Copyright of abstracts refer to PubMed.

Resource Name: GREC Corpus

Resource ID: SCR_006719

Alternate IDs: nif-0000-06688

Record Creation Time: 20220129T080237+0000

Record Last Update: 20260725T190629+0000

Ratings and Alerts

No rating or validation information has been found for GREC Corpus.

No alerts have been found for GREC Corpus.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 3 mentions in open access literature.

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

Vanegas JA, et al. (2015) An Overview of Biomolecular Event Extraction from Scientific Documents. Computational and mathematical methods in medicine, 2015, 571381.

Neves M, et al. (2014) An analysis on the entity annotations in biological corpora. F1000Research, 3, 96.

Thompson P, et al. (2009) Construction of an annotated corpus to support biomedical information extraction. BMC bioinformatics, 10, 349.