

Resource Summary Report

Generated by SPARC SAWG on Aug 9, 2026

Automated recognition of brain region mentions in neuroscience literature.

RRID:SCR_002731

Type: Tool

Proper Citation

Automated recognition of brain region mentions in neuroscience literature.
(RRID:SCR_002731)

Resource Information

URL: <http://www.chibi.ubc.ca/WhiteText/>

Proper Citation: Automated recognition of brain region mentions in neuroscience literature. (RRID:SCR_002731)

Description: Freely available corpus of manually annotated brain region mentions created to facilitate text mining of neuroscience literature. The corpus contains 1,377 abstracts with 18,242 brain region annotations. Interannotator agreement was evaluated for a subset of the documents, and was 90.7% and 96.7% for strict and lenient matching respectively. We observed a large vocabulary of over 6,000 unique brain region terms and 17,000 words. For automatic extraction of brain region mentions we evaluated simple dictionary methods and complex natural language processing techniques. The dictionary methods based on neuroanatomical lexicons recalled 36% of the mentions with 57% precision. The best performance was achieved using a conditional random field (CRF) with a rich feature set. Features were based on morphological, lexical, syntactic and contextual information. The CRF recalled 76% of mentions at 81% precision, by counting partial matches recall and precision increase to 86% and 92% respectively. We suspect a large amount of error is due to coordinating conjunctions, previously unseen words and brain regions of less commonly studied organisms. We found context windows, lemmatization and abbreviation expansion to be the most informative techniques. We encourage you to test new methods and applications of the dataset. Please contact us if you do, we would like to hear about and link to your work. The abstracts are from PubMed/Medline, specifically The Journal of Comparative Neurology.

Abbreviations: WhiteText

Synonyms: WhiteText - annotated neuroscience text

Resource Type: data or information resource, data set

Defining Citation: [PMID:19750194](https://pubmed.ncbi.nlm.nih.gov/19750194/)

Keywords: brain region, annotation, journal of comparative neurology, neuroanatomy, brain

Availability: THIS RESOURCE IS NO LONGER IN SERVICE

Resource Name: Automated recognition of brain region mentions in neuroscience literature.

Resource ID: SCR_002731

Alternate IDs: nif-0000-24027

Alternate URLs: <http://www.nitrc.org/projects/whitetext/>

Record Creation Time: 20220129T080215+0000

Record Last Update: 20260808T190548+0000

Ratings and Alerts

No rating or validation information has been found for Automated recognition of brain region mentions in neuroscience literature..

No alerts have been found for Automated recognition of brain region mentions in neuroscience literature..

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](#) .

French L, et al. (2015) Text mining for neuroanatomy using WhiteText with an updated corpus and a new web application. *Frontiers in neuroinformatics*, 9, 13.

Richardet R, et al. (2015) Large-scale extraction of brain connectivity from the neuroscientific literature. *Bioinformatics (Oxford, England)*, 31(10), 1640.

French L, et al. (2009) Automated recognition of brain region mentions in neuroscience literature. *Frontiers in neuroinformatics*, 3, 29.