

OrganismTagger: Detection, normalization, and grounding of organism entities in biomedical documents

Submitted by [rene](#) [1] on Sun, 2011-08-14 17:00

- [OrganismTagger](#) [2]
- [Bioinformatics](#) [3]
- [NLP](#) [4]

Title	OrganismTagger: Detection, normalization, and grounding of organism entities in biomedical documents
Publication Type	Journal Article
Year of Publication	2011
Authors	Naderi, N. [5], T. Kappler [6], C. J. O. Baker [7], and R. Witte [8]
Refereed Designation	Refereed
Journal	Bioinformatics
Volume	27
Issue	19
Pagination	2721--2729
Date Published	August 9, 2011
ISSN	1460-2059 (online) 1367-4803 (print)
Abstract	<p>Motivation: Semantic tagging of organism mentions in full-text articles is an important part of literature mining and semantic enrichment solutions. Tagged organism mentions also play a pivotal role in disambiguating other entities in a text, such as proteins. A high-precision organism tagging system must be able to detect the numerous forms of organism mentions, including common names as well as the traditional taxonomic groups: genus, species, and strains. In addition, such a system must resolve abbreviations and acronyms, assign the scientific name, and if possible link the detected mention to the NCBI Taxonomy database for further semantic queries and literature navigation.</p> <p>Results: We present the OrganismTagger, a hybrid rule-based/machine learning system to extract organism mentions from the literature. It includes tools for automatically generating lexical and ontological resources from a copy of the NCBI Taxonomy database, thereby facilitating system updates by end-users. Its novel ontology-based resources can also be reused in other semantic mining and linked data tasks. Each detected organism mention is normalized to a canonical name through the resolution of acronyms and abbreviations and subsequently grounded with an NCBI Taxonomy database ID. In particular, our system combines a novel machine-learning approach with rule-based and lexical methods for detecting strain mentions in documents. On our manually annotated OT corpus, the OrganismTagger achieves a precision of 95%, a recall of 94% and a grounding accuracy of 97.5%. On the manually annotated corpus of Linnaeus-100, the results show a precision of 99%, recall of 97% and grounding accuracy of 97.4%.</p>

Availability: The OrganismTagger, including supporting tools, resources, training data and manual annotations, as well as end-user and developer documentation, is freely available under an open source license at

<http://www.semanticsoftware.info/organism-tagger> [9].

URL

<http://bioinformatics.oxfordjournals.org/cgi/content/abstract/btr452?ikey=9uBtAIDN8y2nmz6&keytype=ref> [10]

DOI

[10.1093/bioinformatics/btr452](http://dx.doi.org/10.1093/bioinformatics/btr452) [11]

Impact Factor

5.468 (2012)

History

Received on March 7, 2011; revised on July 14, 2011; accepted on July 31, 2011.

Attachment

[Bioinformatics-2011-Naderi-2721-9.pdf](#) [12]

Size

424.64 KB



Except where otherwise noted, all original content on this site is copyright by its author and licensed under a [Creative Commons Attribution-Share Alike 2.5 Canada License](#).

Source URL (retrieved on 2025-12-22 11:35): <https://www.semanticsoftware.info/biblio/organismtagger-bioinformatics2011>

Links:

[1] <https://www.semanticsoftware.info/users/rene>

[2] <https://www.semanticsoftware.info/taxonomy/term/249>

[3] <https://www.semanticsoftware.info/category/topic/bioinformatics>

[4] <https://www.semanticsoftware.info/category/topic/nlp>

[5] <https://www.semanticsoftware.info/biblio/author/22>

[6] <https://www.semanticsoftware.info/biblio/author/30>

[7] <https://www.semanticsoftware.info/biblio/author/16>

[8] <https://www.semanticsoftware.info/biblio/author/1>

[9] <http://www.semanticsoftware.info/organism-tagger>

[10] <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/btr452?ikey=9uBtAIDN8y2nmz6&keytype=ref>

[11] <http://dx.doi.org/10.1093/bioinformatics/btr452>

[12] <https://www.semanticsoftware.info/system/files/Bioinformatics-2011-Naderi-2721-9.pdf>