protocols.io

drNER Version 2

Tome Eftimov, Barbara Koroušić Seljak, Peter Korošec

Abstract

Test cropora for evaluation of drNER, together with the obtained results for each document, separetly.

Guidelines

drNER is a rule-based Named-entity recognition method for knowledge extraction of evidence-based dietary recommendations.

The files are available at: http://cs.ijs.si/eftimov/drNER.

The "Scientifically_validated_documents.xlsx" contains 50 documents extracted from scientifically validated web sites that are used for evaluation of drNER.

The "PubMED_abstracts.xlsx" contains 50 documents that are PubMED abstracts and they are used for evaluation of drNER.

In the "drNER results" folder, the obtained results by the drNER are presented. For each document the obtained results are presented into "csv" file, "DocNumber.csv". Number that is presented in the name of the "csv" file is related to a document from the test sets. The documents from "Scientifically_validated_documents.xlsx" have numbers form 1 to 50, while the documents from "PubMED_abstracts.xlsx" have numbers from 51 to 100.

Each "csv" file consists of 6 parameters, "Entity", "Label", "SOP label", "Sentence segment", "Sentence", and "Document". "Entity" is the extracted phrase from the document. "Label" is the category to which the extracted entity belongs. "SOP label" gives indication about the location from where the entity is extracted (S- Subject, P - Predicate, O - Object). "Sentence segment" is the number of sentence segment if the sentence is split into segments according to the preprocessing

that is defined. "Sentence" corresponds to the number of sentence when the text is split on sentences. "Document" is the number of document from the test set.

CONTACT

tome.eftimov@ijs.si

Protocol

Step 1.