We have developed a python module and a web application, PPaxe, that allows users to extract PPIs and protein occurrence from a given set of PubMed and PubMed Central articles, based on abstracts and full-texts respectively. PPaxe tokenizes and annotates the sentences with StanfordCoreNLP [2] and then distills a number of features that are analyzed by a Random Forest classifier. Finally, it presents the results of the analysis in different ways to help researchers export, filter and analyze the interactions easily.

Acknowledgements

This work was supported by Spanish Ministry of Economy (BFU2014-56055P), Generalitat de Catalunya (2014SGR687, 2018SGR1455); Predoctoral fellowship by AGAUR (FI-FDR, 2017FI_B_00191).

References