EXTRACT

Literature Mining vLab – LifeWatch Greece

Lifewatch Greece Portal

E-mail Password Sign In

Forgot your password? Register

Literature Mining

https://lm.portal.lifewatchgreece.eu

Biodiversity literature and data constitute a vast public resource open to mining and knowledge extraction.

Associating organisms to key features of their life, for example the environment in which they live, the way feed, their breeding habits, is cornerstone in explaining biodiversity patterns and informing ecological decisions.

The Literature Mining virtual Lab (LM-vLab) aims at both:

- the automatic extraction of species - traits associations from the literature
- augmenting Lifewatch Greece species related information based on the above

The EXTRACT annotation tool, and the ENVIRONMENTS and SPECIES/ORGANISMS taggers are relevant LM-vLab tools to this end. All three are being employed for standard compliant term suggestion to describe the environmental context of metagenomic records, while ENVIRONMENTS is also being used identification of Environment Ontology terms in text and the annotation of the Encyclopedia of Life.

All tools are developed in collaboration and maintained at the group of Prof. Lars Juhl Jensen, Novo Nordisk Foundation Center for Protein Research, Copenhagen, Denmark.

Developed by HCMR and FORTH

http://extract.hcmr.gr

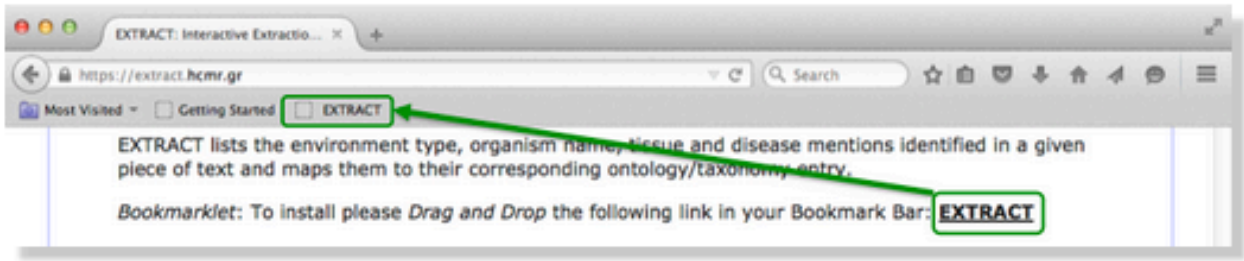
EXTRACT
Interactive Extraction of Metadata

extract@hcmr.gr

About | Demo | Help

EXTRACT lists the environment type, organism name, tissue and disease mentions identified in a given piece of text and maps them to their corresponding ontology/taxonomy entries.

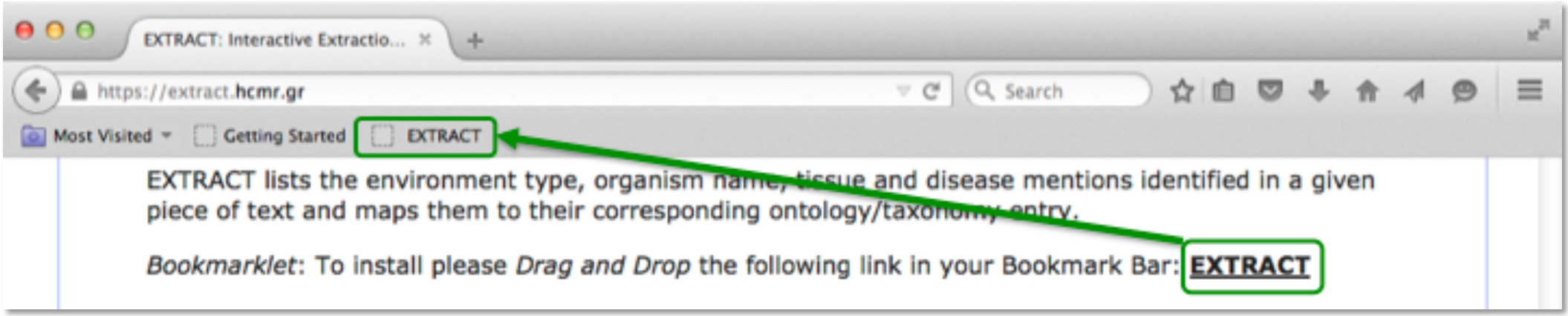
Bookmarklet: To install please *Drag and Drop* the following link in your Bookmark Bar: EXTRACT



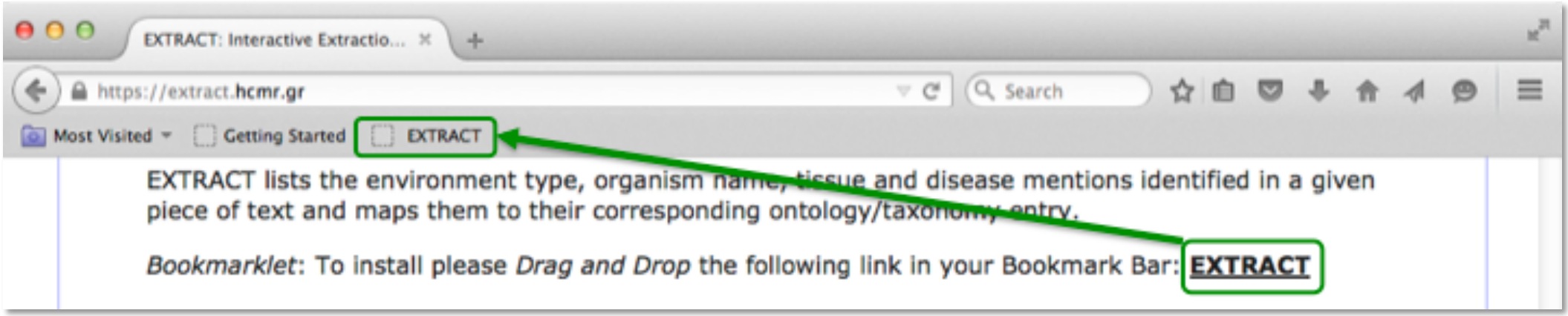
Usage: a. *select* a piece of text of interest in a web page and then b. *click on the bookmarklet*. c. A *pop-up* such as the following will appear (supported browsers: Chrome, Firefox, Safari). By *hovering* the mouse cursor over the text tags or the table rows you can visually inspect which words have been identified as which entities.

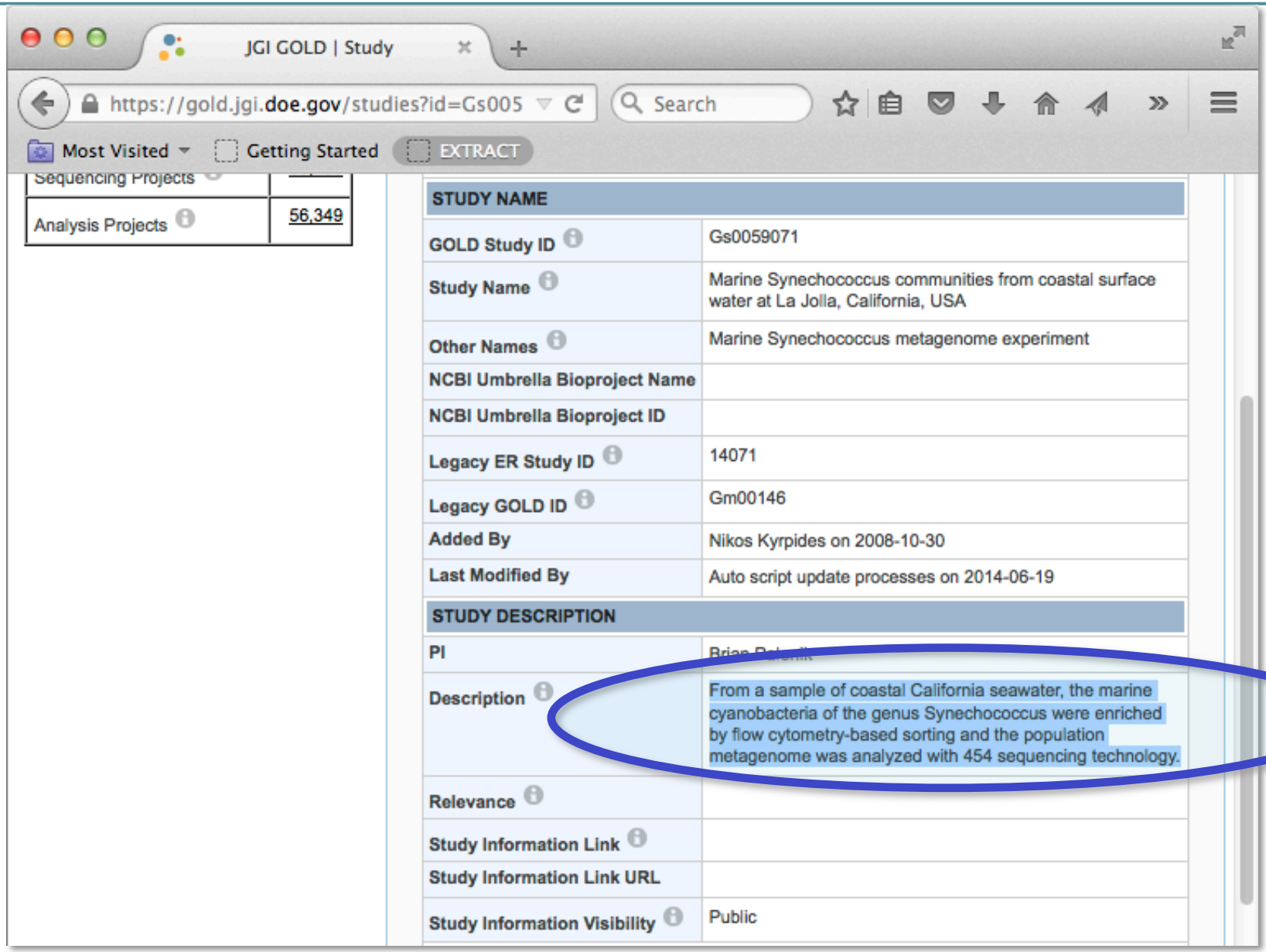
EXTRACT: interactive extraction of environment metadata and term suggestion for metagenomic sample annotation Pafilis E, Buttigieg PL, Ferrell B, et al.. (2016). **Bioinformatics, 2016**, baw005. doi:10.1093/bioinformatics/btv04

Drag n' Drop Installation

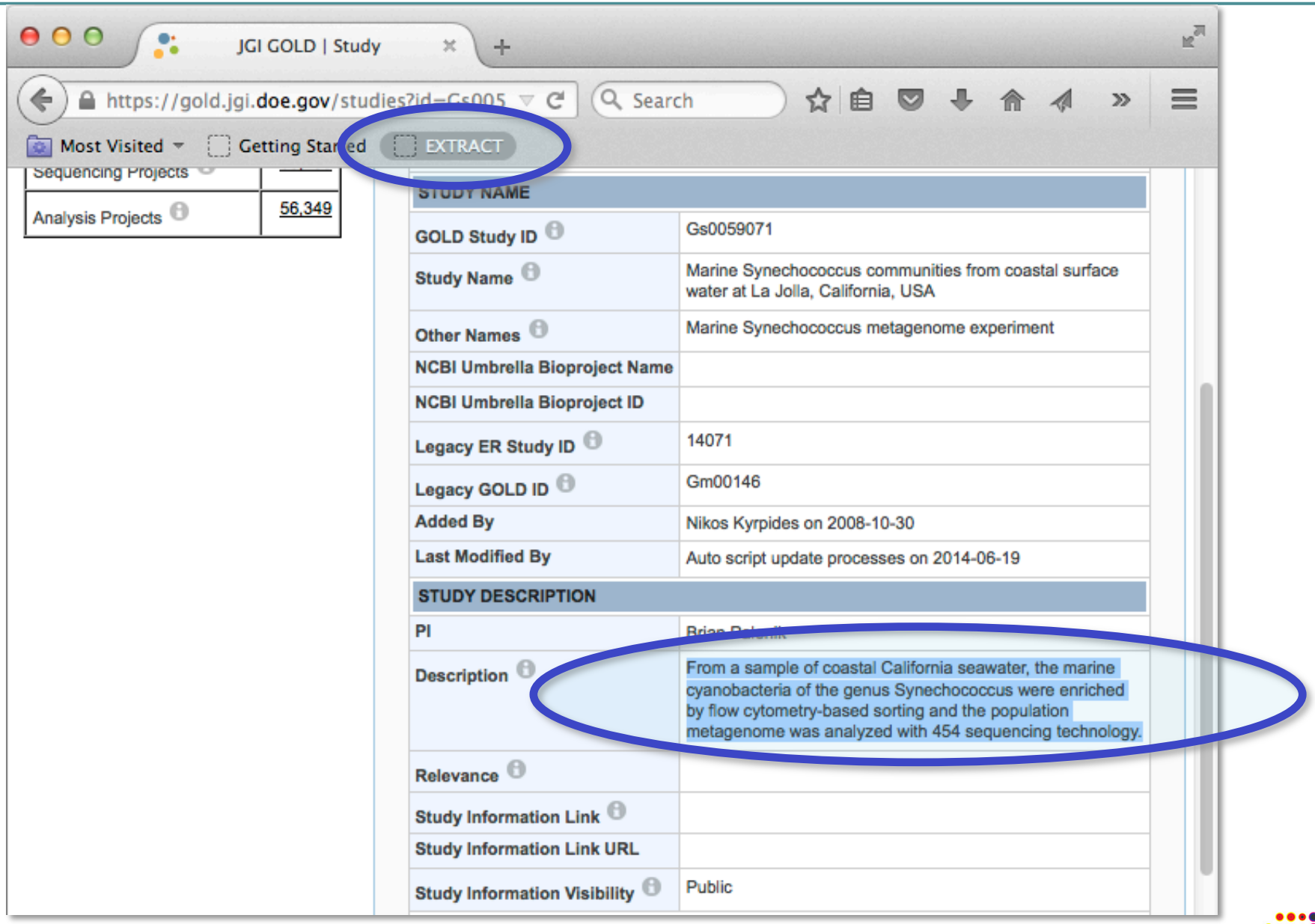


Drag n' Drop Installation



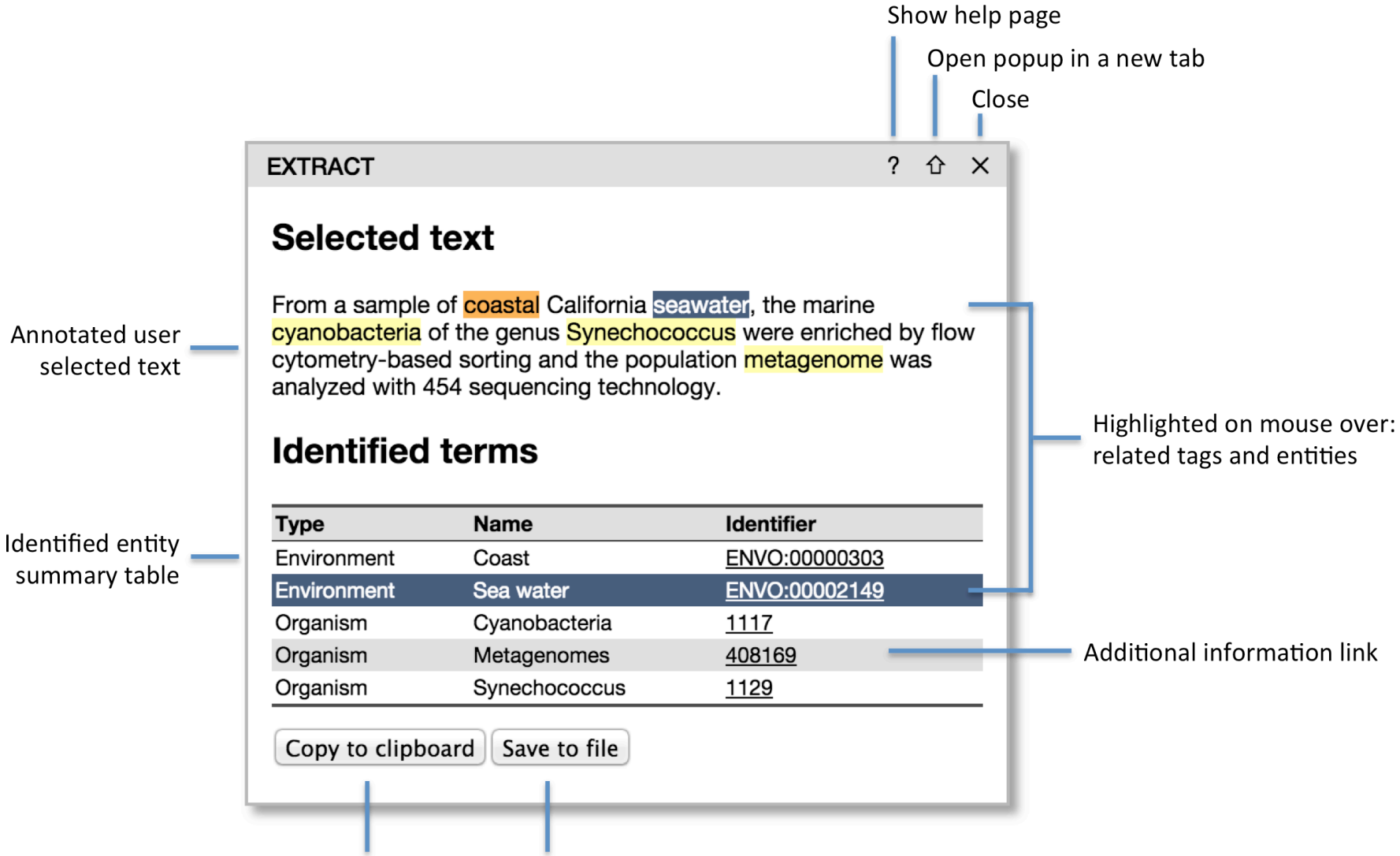


https://gold.jgi.doe.gov/studies?id=Gs0059071

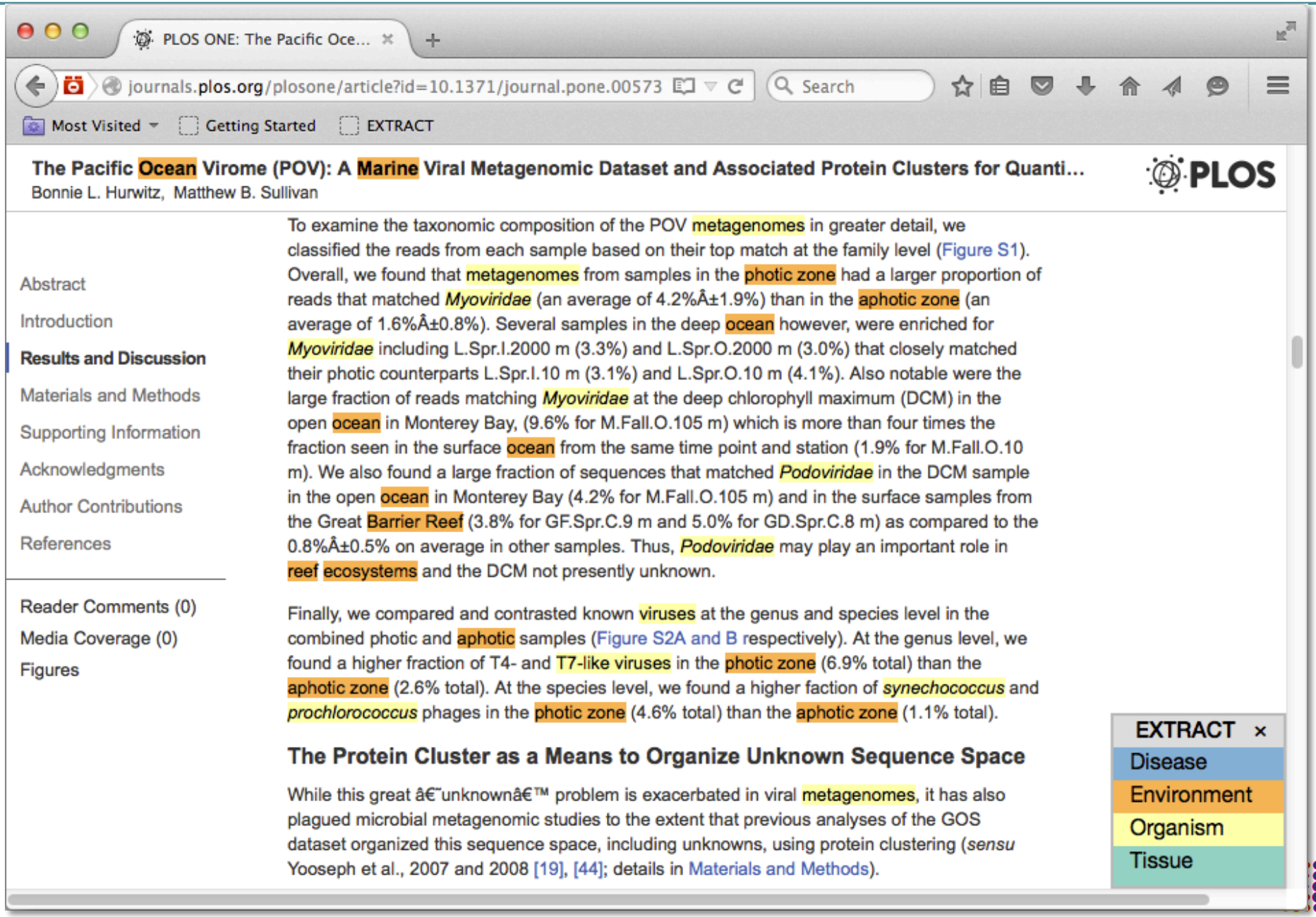


https://gold.jgi.doe.gov/studies?id=Gs0059071

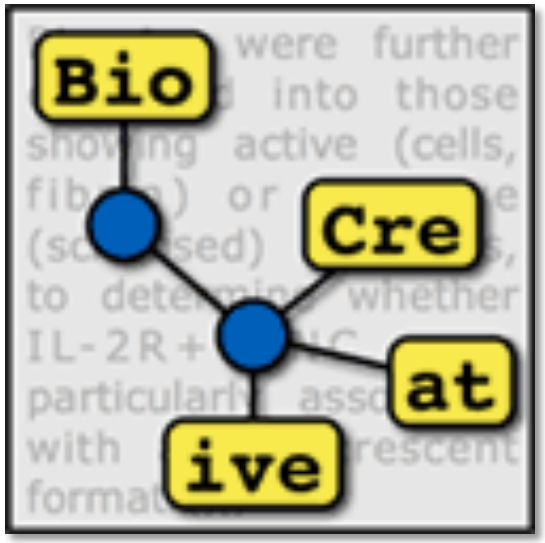
Popup – Multiple Entities



Full Page Tagging



Hurwitz & Sullivan, PONE, 2013, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0057355#s2



http://www.biocreative2015.org

BioCreative V: Interactive Annotation Task (IAT) Dr. L. Hirschman, Dr. C. Arighi *et al.*

Challenge: **March – August 2015**

Presentation**: September 2015**, Sevilla, Spain

Metagenomics Record Annotation Session

(Department of Energy [DE-SC0010838])

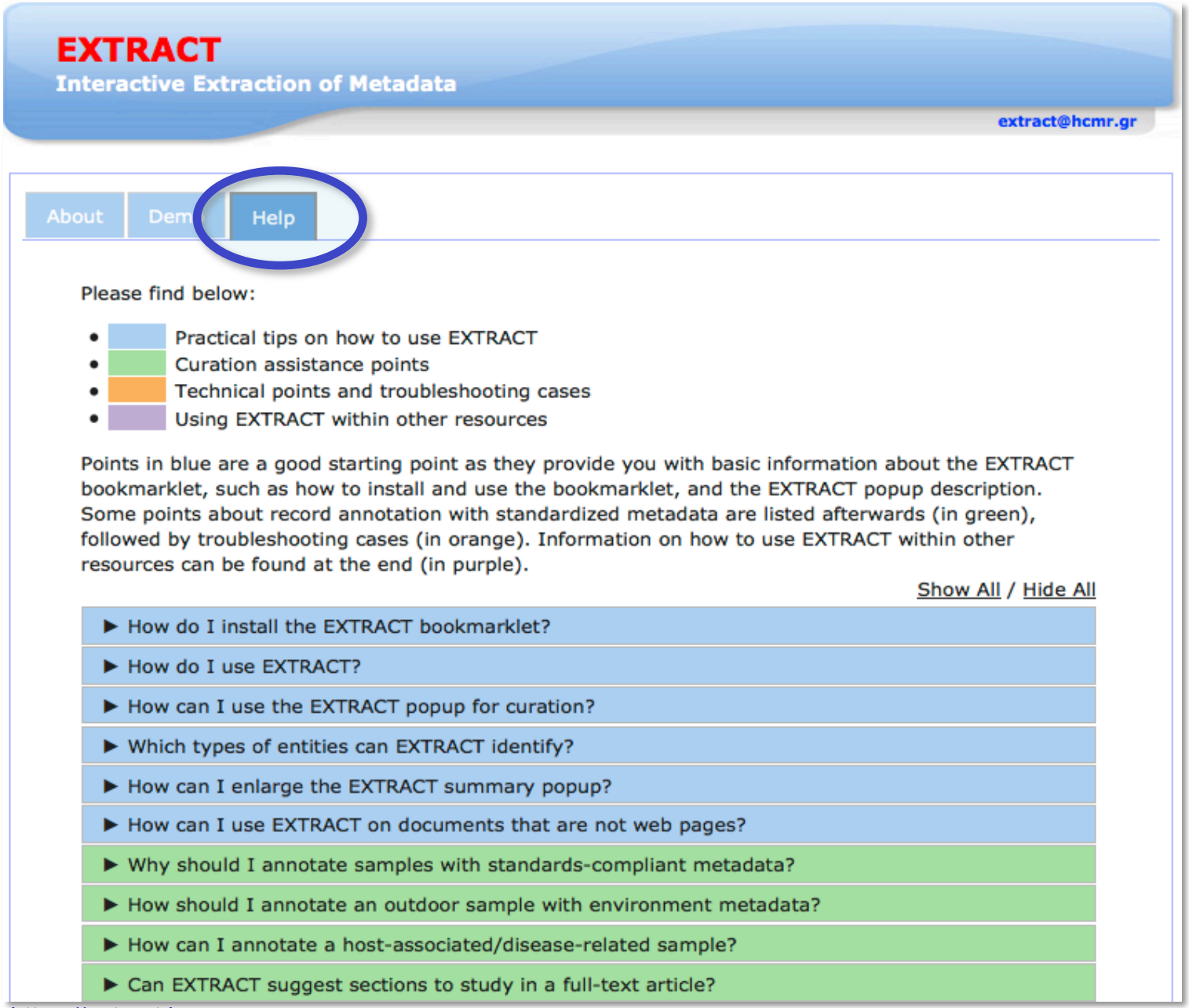
Pafilis E, Buttigieg PL, Ferrell B, *et al*. (2015). Proceedings of the Fifth BioCreative Challenge Evaluation Workshop, 384–395.
http://www.biocreative.org/media/store/files/2015/IAT_extract_1.pdf

EXTRACT: interactive extraction of environment metadata and term suggestion for metagenomic sample annotation Pafilis E, Buttigieg PL, Ferrell B, et al.. (2016). **Bioinformatics, 2016**, baw005. doi:10.1093/bioinformatics/btv04

Metagenomics Expert Evaluation

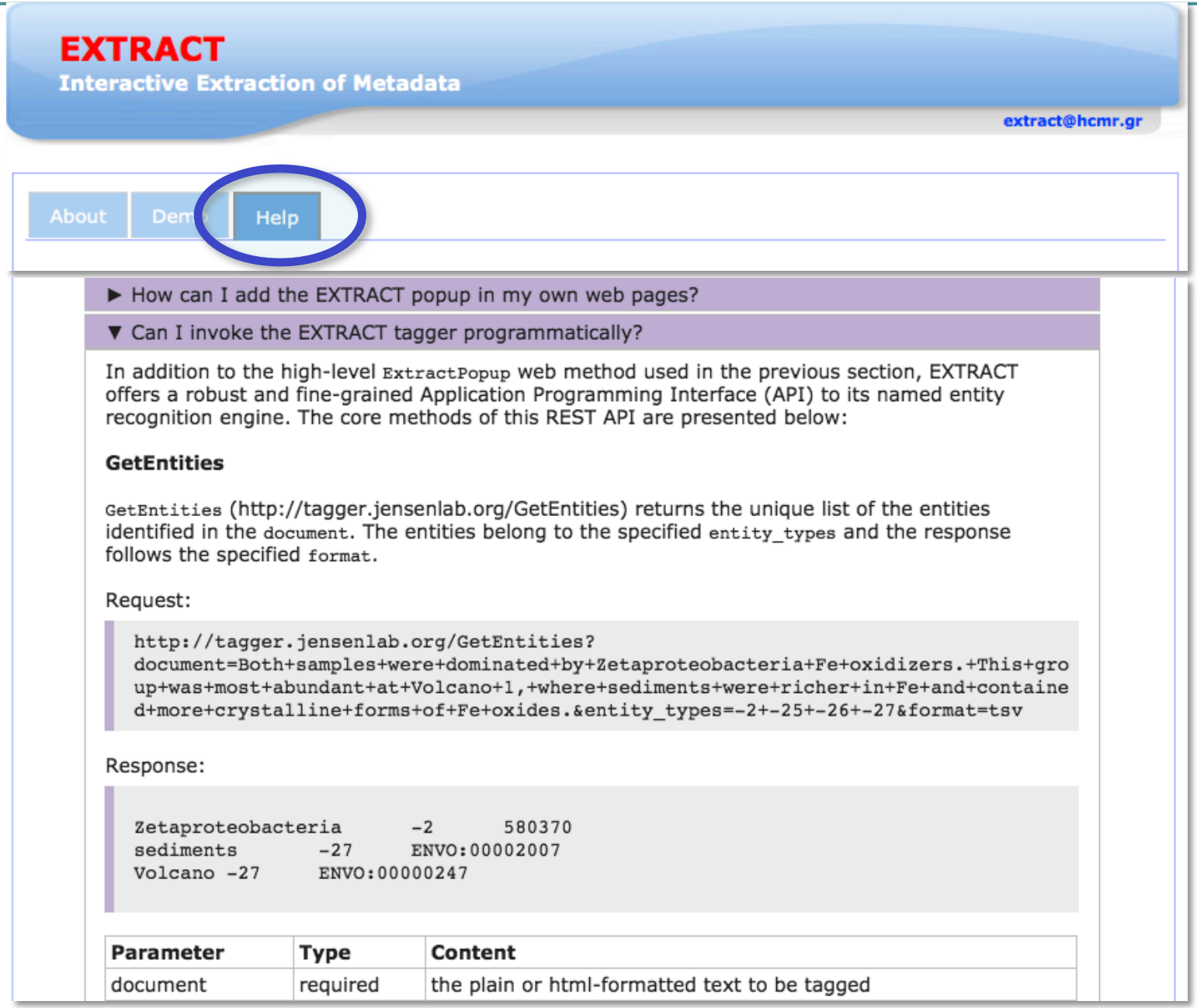
- Easy: installation, tagging a web page, invoke the popup by processing selected text, saving results to a file
- Adequate NER accuracy (4/10 reported FN/FP, still two of them satisfied)
- Speedup in the range of 15–25%.
- Saving by time by avoiding looking up the ENVO identifier for every term
- Manual document inspection still needed
- EXTRACT helps in finding terms that would have been missed by the curators (e.g. due to non-familiarity with terminology)
- Average score 8.3 out of 10: they would recommend EXTRACT

Documentation



https://extract.hcmr.gr

API Documentation



https://extract.hcmr.gr, screenshot cropped for illustration purposes

Thank You