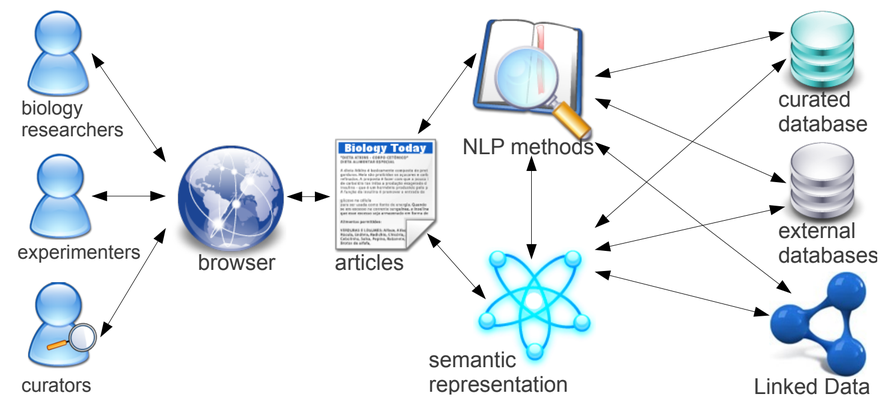
Proxiris, a Browser Companion for Biocurators of Fungal Genomics Literature

Vahé Chahinian, Marie-Jean Meurs, David Mason, Erin McDonnell, Ingo Morgenstern, Greg Butler, Adrian Tsang

Centre for Structural and Functional Genomics, Concordia University, Canada

Literature Mining and Curation

- Reading, interpreting, curating bio-literature
- Labor-intensive, error-prone and expensive task
  - ⇒ Natural Language Processing (NLP) techniques:
    - ✓extract knowledge from papers
  - ⇒ Semantic techniques:
    - ✓connect information from various sources



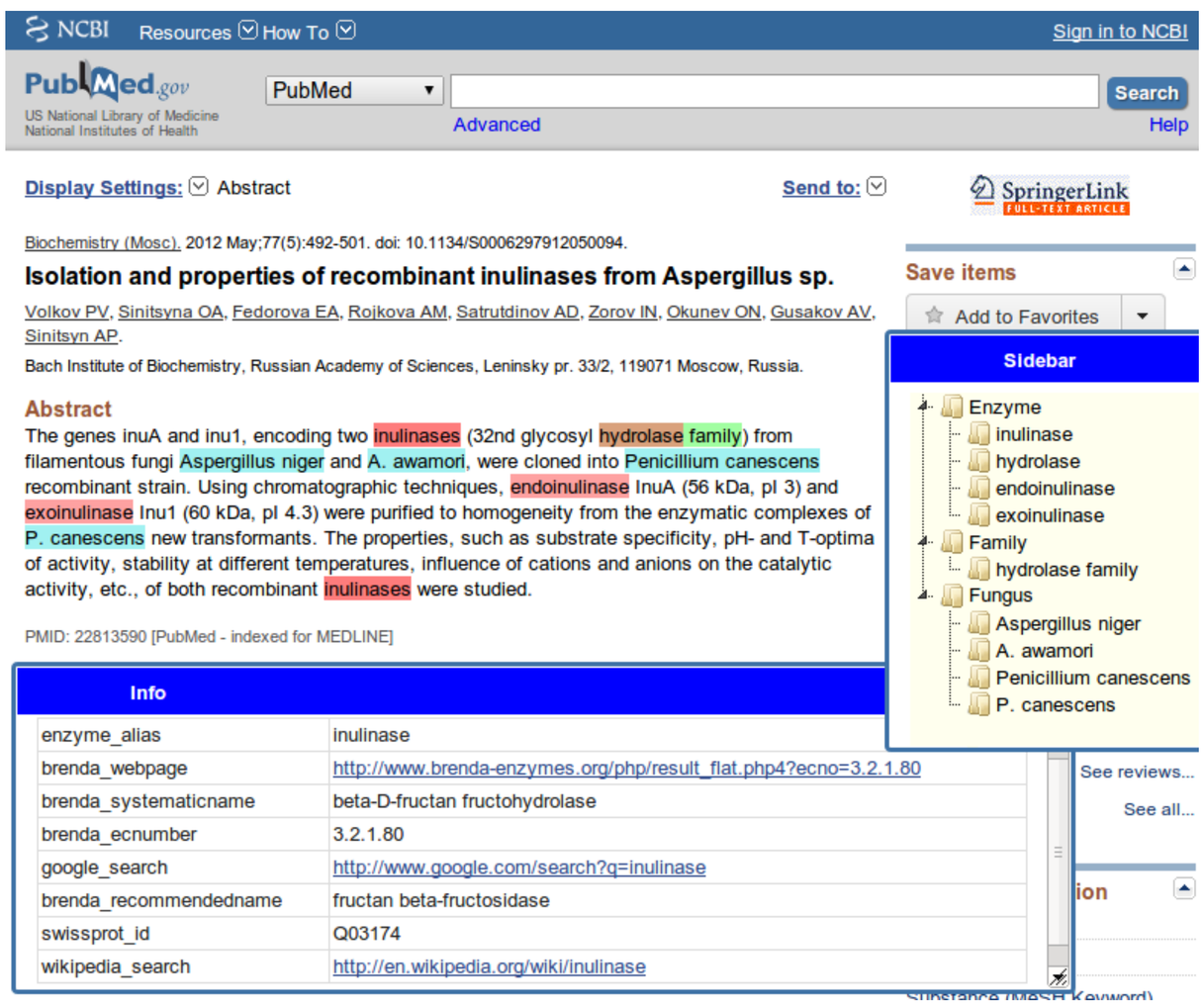
Proxiris > what you see

- Augmented browsing for literature curation
- Support genomics-based lignocellulose research [2]
- Content processing with mycoMINE [1]
- Highlight facts and entities of interest
- Provide additional content from external DBs:

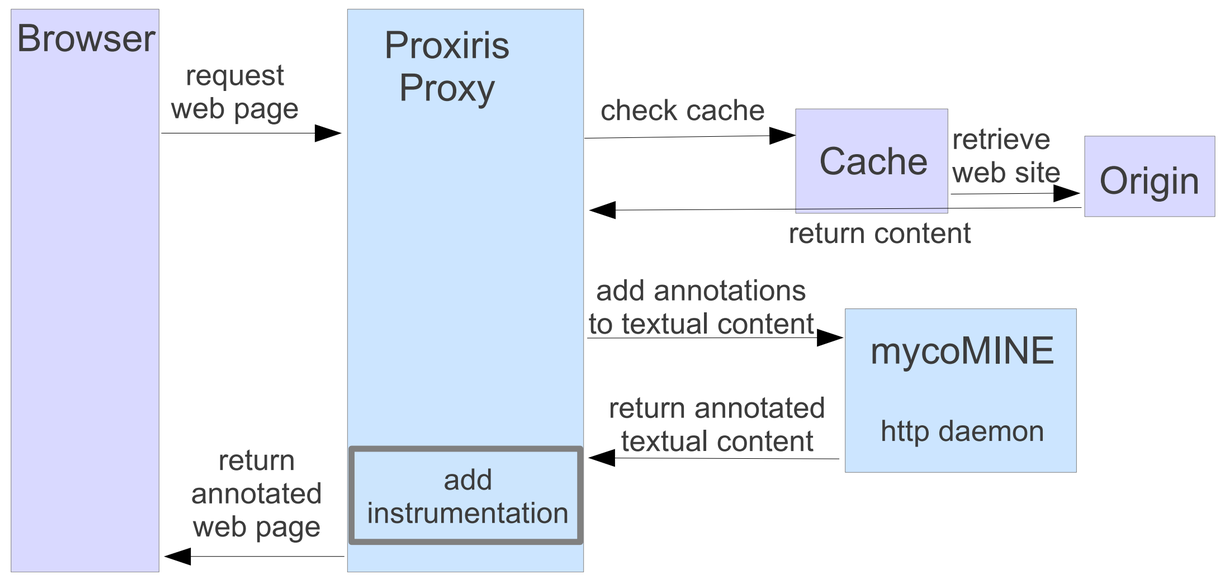
BRENDA identifier, recommended and systematic name, SwissProt identifier, NCBI Taxonomy identifier, CAZy family

- Link to external knowledge sources:

BRENDA, SwissProt, NCBI, Google, Wikipedia.



Proxiris > how it works



Implementation

- User's browser uses Proxiris Proxy server
- Server retrieves requested pages
- Node.js server on the proxy sends HTML parts to mycoMINE
- Original content is replaced with processed text
- Proxiris instrumentation = JavaScript [jQuery, node.js, jsTree] and CSS
- Node.js processes returned HTML documents to support highlighting functionality
- Node.js injects Proxiris JavaScript, jQuery and CSS into the original webpage
- JavaScript and jQuery implement all of the dynamic features (left click, highlight toggle sidebar functionality)
- JsTree builds the sidebar
- CSS provides all the HTML markup descriptions

A successfull approach > how and why

- Evaluation: 2 curators, 114 PubMed abstracts, Proxiris prototype with limited features ⇒ triage time reduced by 21%

Proxy approach:

- flexible and generic ► easy integration of text mining and semantic services
- browser independance ► no limitation in browser selection, no custom code
- preserves format of original documents ► pictures, tables, embedded services protected
- circumvents same origin policy
- systematic processing of publications from selected origins ► quickly available for users
- service available only on selected web sites ► high level quality of service

Acknowledgment

Funding: Genome Canada, Génome Québec
Collaboration: Justin Powlowski
Technical support: Andrei Wasylyk

References

[1] Meurs et al., *Semantic text mining support for lignocellulose research*, BMC MIDM, 2012

[2] Murphy et al., *Curation of characterized glycoside hydrolases of fungal origin*, Database, 2011