To infinity and beyond….

TEXTPRESSO

What is it?

- Text-mining system for scientific literature
  - Full body text search
  - Biological categories
    - Groups of terms describing a concept
  - personal curation
- Developed by Hans-Michael Muller et al.
  - part of WormBase at CalTech
  - Extensive C. elegans curation
  - *Textpresso: an ontology-based information retrieval and extraction system for biological literature*
    Muller HM, Kenny EE, Sternberg PW
    PLoS Biol. 2004 Nov;2(11):e309. Epub 2004 Sep 21.
- Textpresso.org

Current Status

- Fork project
  - Hosted on Sourceforge
    - http://sourceforge.net/projects/textpresso
- Interface overhaul
  - Heavy integration with jquery datatables
    - http://datatables.net
  - User login/registration
  - Manage corpus through GUI
- Significant code rewrite
  - Modularization/customization
    - Better database support
    - Plug-in API
  - Ajaxification
  - Speed increase
    - code works poorly with MacOSX

Future Roadmap

- Web services
- Smarter paper analysis
  - Methods tagging
  - Better PageRank
    - references
- In-browser paper display/highlighting
  - a la Google Books
- Better command line functionality
- Proper GMOD integration
  - Bioperl mashups

Questions/Suggestions?

- nml5566@gmail.com