OpenRefine

Phase 2 enhancements

foundation@openrefine.org

Overview

This proposal begins to summarize Phase 2 efforts to enhance OpenRefine with more features, enhancements, and changes requested from our community and in particular News Organizations.

Financial Summary

Accounts Payable / Receivable at this link: **TBD**

Goals for Phase 2 Enhancement Funding

With this 2nd proposal, we plan to address the tight coupling between the front and back end. The current 9 year old implementation presents technical barriers and limits new features & new contributors to OpenRefine.

Front / Backend separation

- ○ Scope: completely separating the backend so that a full API can be exposed for all OpenRefine operations and commands. Once done, we can move to a modern front-end framework and support a different back-end engine.
- ○ Deliverable: Functional and documented API covering all the commands available in OpenRefine 3 front end. See #1483

Other Long term Goals

Items below are listed in no specific order.

UX Refresh and Libraries

- Scope: Leverage the new API and Datagrid to move away from Jquery-based components and using React or Angular.
- DEPENDENCIES: Front / Backend separation required

Datagrid Enhancements

- Scope: Faster grid technology with virtual dom for faster scrolling, and interactive elements like drag and drop column moving and resizing, and column and row grouping to better support record mode.
  i. Some parts of the front end are highly coupled with the back end. Using a newer datagrid will help remove those items.
  ii. See [#1347](#1347)

Extension Support 2.0

- Scope: Ensure that we have extendable points for our most important components in OpenRefine that the community could conceivably want to extend.
- (Financially) Help extension maintainers to migrate their project to the new OpenRefine API and front end (defined during Phase 2)

Support Modern Data Processing Engine

- Thanks to the decoupling of the front and back, we can answer the numerous request from the community to support new data processing engine like Apache Spark or Google DataFlow / DataProc.
- See GitHub project [Performance Improvements](#)

Data Storage Enhancements

- Scope: Increase the capacity of OpenRefine's data processing by moving to an alternative, higher performance in-memory / on-disk data storage technology.
- See [#1469](#1469)

R Lang support

- There is significant use of statistics within News Organizations where the goal of minimizing the back and forth between R tooling and OpenRefine would be explored and assessed by the community.
- Once defined goals are laid out via Community Events, Mailing List, etc. then features would be added to our backlog and defined inside Development Milestones in later sections of this document.
- See #1226 and wiki page OpenRefine integration ideas with R lang and Jupyter
- DEPENDENCIES: UX REFRESH required

Planned Support Activities

Planned Event Attendances:
- To Be Determined

General Support Efforts:
- To Be Determined