Kaiāulu

Anthony Lau, Ian Jaymes Iwata, Sean Sunoo

Sponsored by: Carlos Paradis and Rick Kazman

Technical Solution Overview

- What is Kaiaulu
  - Kaiaulu is open source: https://github.com/sailuh/kaiaulu
  - Data miner and analysis
  - R package that consists of notebooks, functions, config files
  - Written as part of a PhD thesis in UH Manoa
- Problem
  - No way to automatically update downloaded data
  - Need to manually download any new data
  - No overarching structure for file paths or config file organization
- Solution
  - Create functions that parse most recent data from already present files
  - Unify file path organization and universal file naming conventions
  - Create functions that call various REST API to retrieve new data
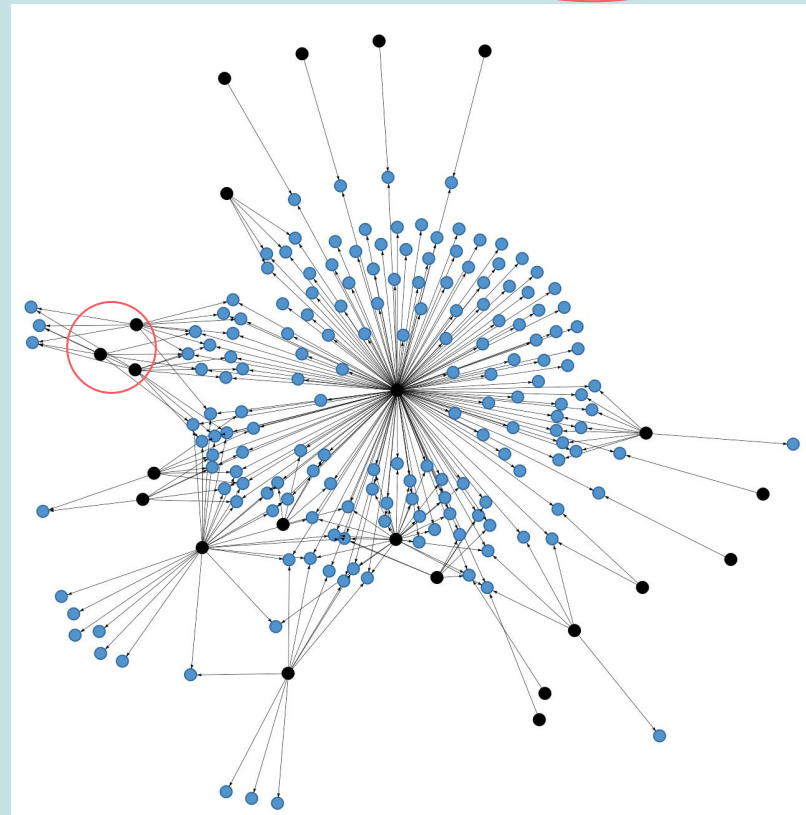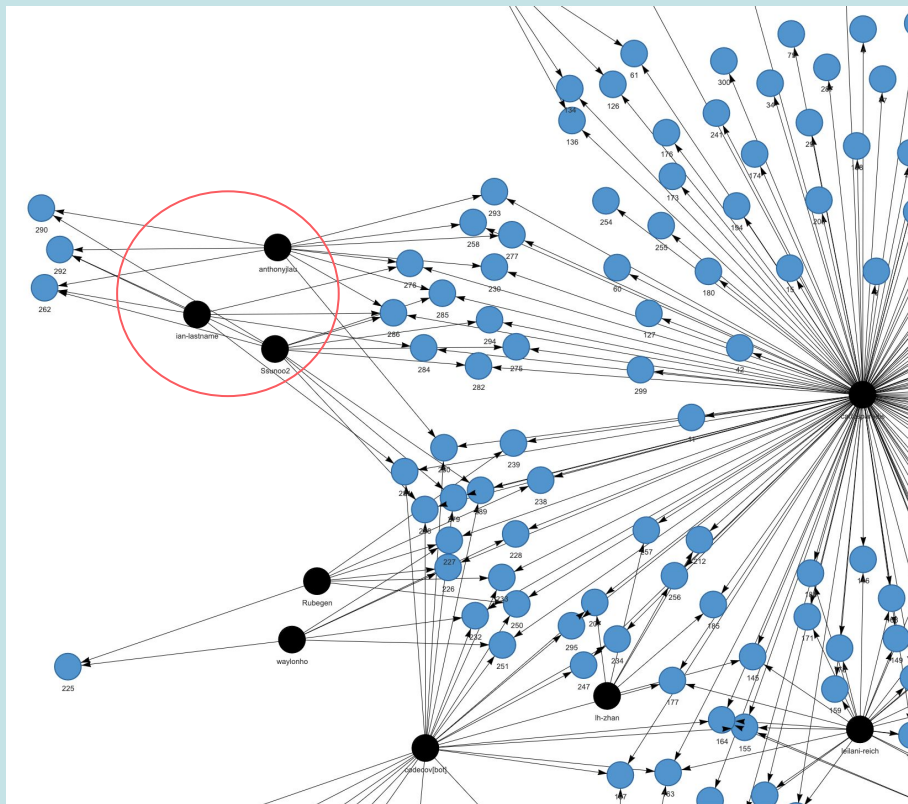 - Write scripts which can be executed from the command line by Cron Jobs
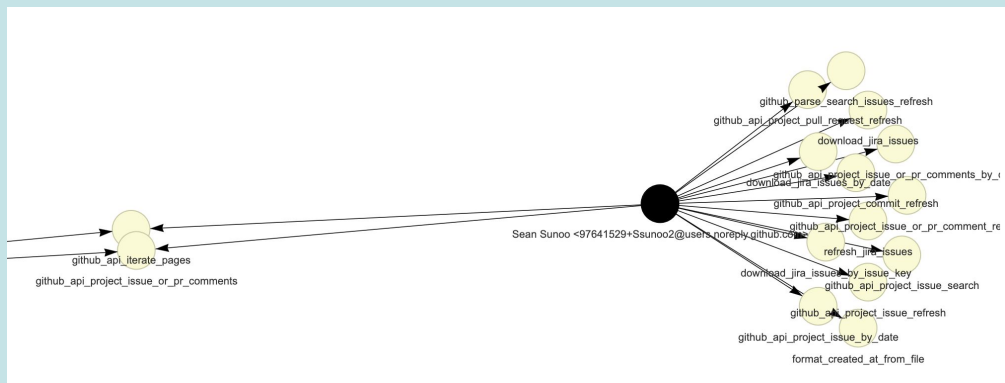
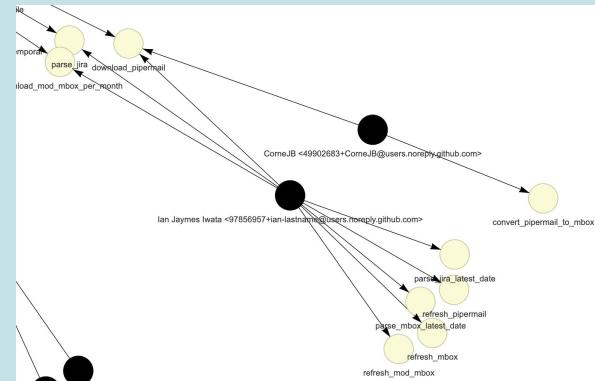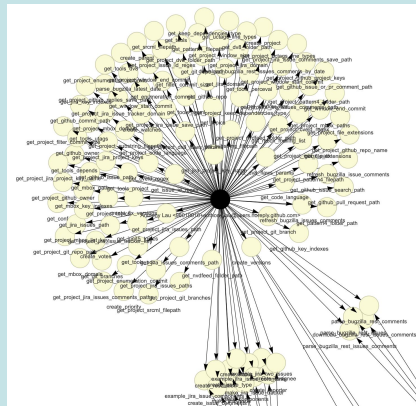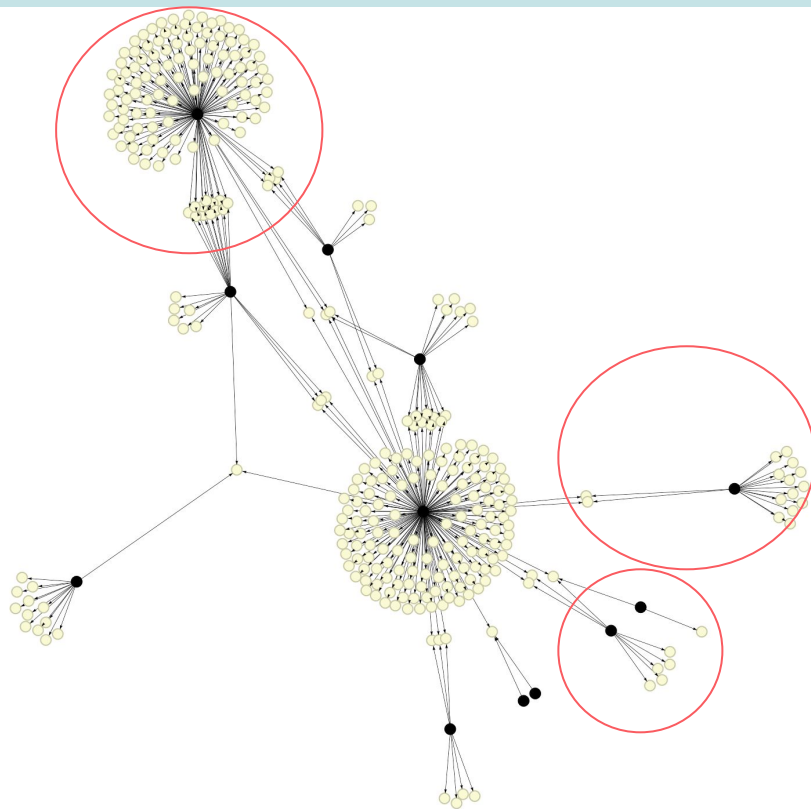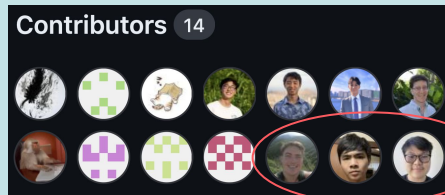Kaiāulu

Major Accomplishments

- Improved existing functionalities
- Config files/ naming conventions
- Refresher functions
- Scripts for automatic downloading via Cron jobs
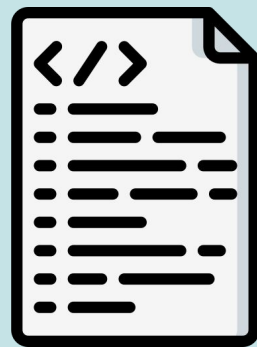
# Major Accomplishments

Major Accomplishments

Improve Existing Functionalities

- JIRA parser
  - Parses all json files in a folder in correct format
  - Added previously missing fields

- JIRA fake data generator
  - Used to unit test downloader's parsers
  - Removed use of third party dependency
  - Formats fake json file in same way as API calls
  - Adheres to format downloaded from API
    - Important to not modify data downloaded
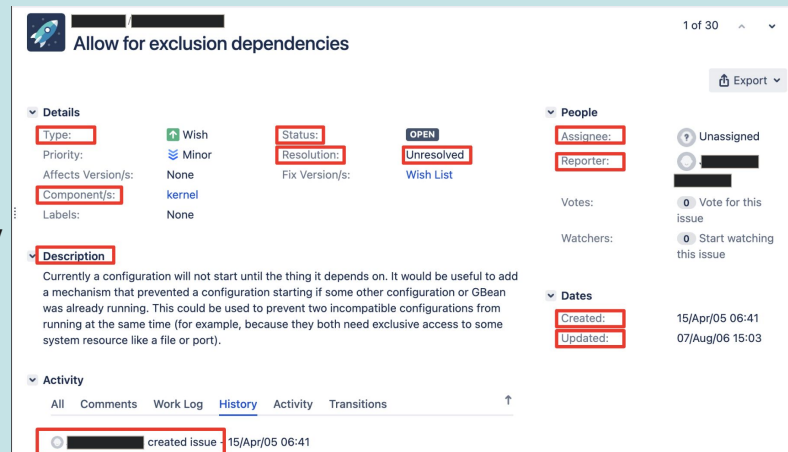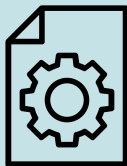
Config Files/Naming convention

- Naming Convention
  - identifier_[UNIX start time]_[UNIX end time].json
  - Allows identification of file containing latest data by only parsing file names
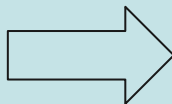
- Config File Get Functions
  - Get functions for config file values
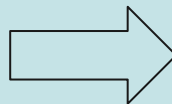 - Easier to propagate changes in config files to notebooks

kaiaulu_13625_32473.json
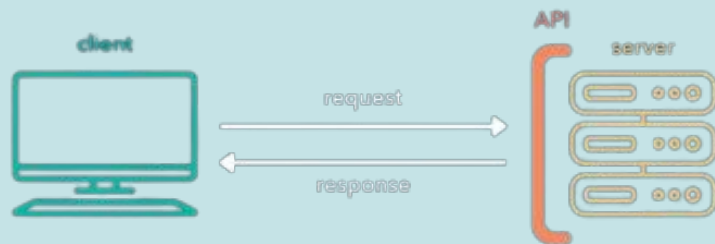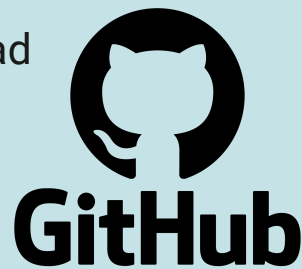
Config File → Get Functions → Kaiaulu Notebooks

Refreshers

- Solution to obtain new issues/comments posted since last download
    - 1. The rest API rate limit is reached.
    - 2. Function ends before completing.
    - 3. More issues were since posted after downloading.

- Refresh functionalities for various endpoints
    - Github
    - JIRA
    - Bugzilla
    - Mail Archives

Scripts/Monitoring

- Wrapper scripts callable from the command line
- Allows for automatic retrieval of new data in time increments
- Cron jobs call scripts every x minutes
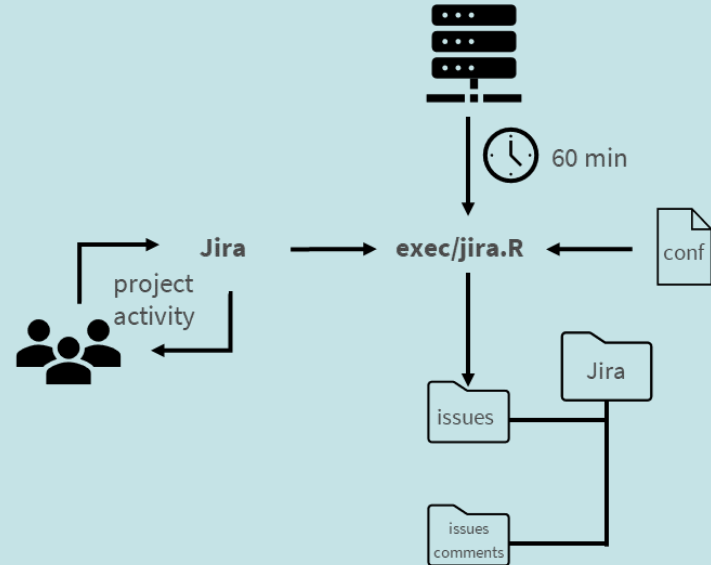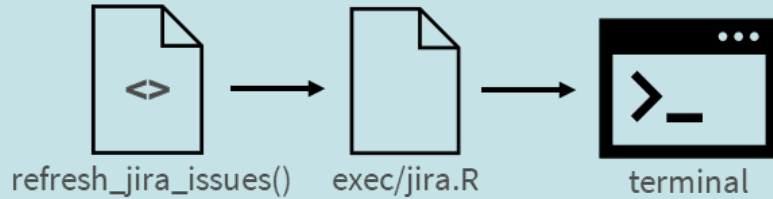
Quality Assurance Metrics

- Unit Testing
  - Test input and output of functions
  - Ensure that config files provide correct information

- End-to-end Testing
  - Check that correct data downloaded
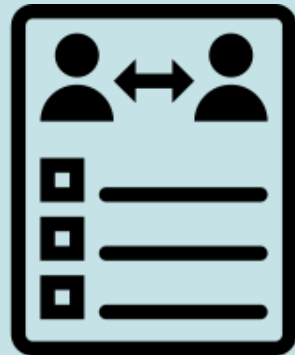 - Parse existing data into tables

Project Management and Delivery

- Project Management
  - GitHub Issues for conceptual discussion of problem and solutions
  - GitHub Discussions for user issues and discussion
  - GitHub code reviews
  - GitHub repo for meeting notes and agenda
- Communication
  - Discord server with sponsor
  - Weekly meetings
    - Monday/Wednesday 5pm - students
    - Friday 3pm - students + sponsor

Reflection

- Challenges
  - APIs behaved differently than expected
    - Led to evolving milestone requirements
    - 'Search' APIs have limitations
- Technical/Soft skills
  - R/Rstudio
  - Familiarity using different API endpoints
  - GitHub issues and pull requests
  - Regular communication and feedback
  - Navigating open source software on GitHub

Next Steps

Kaiaulu is ever evolving
- Unfinished tasks:
  - GitHub events
  - Full functionality of Bugzilla parser
  - Specify month in endpoint for mail downloaders/refreshers
- For future contributors:
  - Add more refreshers
  - Automate more downloaders via scripts
  - Replace more third-party dependencies with own code
  - Navigate API endpoints and their capabilities/limitations

Acknowledgements

We would like to thank our incredible sponsors/stakeholders:
- Carlos Paradis, Ph.D. in Computer Science
- Rick Kazman, Shidler College of Business