Beagle - Navigating Academic Research

Richard Littauer
richard.littauer@gmail.com

ABSTRACT

Beagle is a client-side application that facilitates both annotation on PDFs and web pages and traversing open source academic research. Structurally, Beagle is envisioned as a browserified node application based on a peer-to-peer distributed file system, accessed through a browser extension and a stand-alone application. The UI is envisioned primarily as inline annotations combined with an informative sidebar overlayed on the PDF or wbe page. The application is inherently modular, which serves the dual purpose of maximising extensibility for the code base (which could then be utilized by other members of the open research community) and allowing users to select what information they can see. This document outlines the technical specifications underlying Beagle.

1. INTRODUCTION

[Motivate Beagle. Introduce problems. Open Research. Citation graph. Competitors. Partners. Chrome Extension. Data storage. User interaction. Encrytion. Future]

2. FRONT ENDS

2.1 Chrome Extension

Beagle is mainly realised as a Chrome extension. The extension loads a browserify bundled javascript file. This file is the concatination of the the node.js modules which are specified by the user; it also contains the React templates used to automatically generate a siderbar and various modals which make up the UI for the extension.

2.1.1 Components

- Author Profile
- Publication Graph
- Tags
- Citation
- Saved papers
- Annotations
- Journal information
- Paper information
- Notes

2.2 Standalone App

Beagle will be able to pass on PDFs and web snippets via email to other users. It will also be able to store PDFs and display them to users who share them with other users.

3. DATA STORAGE

Beagle will store publicly citation data; data about viewed articles; data about authors; publication graph data. It will also have a node available for accessing and storing encrypted user data and annotations in a peer-to-peer distributed file system, allowing for maximal uptime and decentralization of storage costs. Users can share directly via their own node with other uses, bypassing the central storage, if they wish, although all data sent will be hashed.

Some data will be optionally encrypted. For instance, tags on paper will have the option of being public, semi-private (available to user groups), or private. This allows Beagle to use tags to deduce contents of papers, or related papers.

Encrypted data stored on a p2p network instead of in a data silo has several distinct advantages; most namely, that the user is not dependant on a silo for continued access to their data. The ensures the uptime of Beagle, regardless of whether it is under active development or maintainence.

4. USER AUTHENTICATION

Users will register and sign on to the platform using the same process: two step authentication. This is a more secure process than using a password, either server-side or client-side. This has an added benefit of allowing storage of a user's information before they sign on to the site.

When a user loads the extension or the web app, they will be allowed to access some functionality without needing to log in. For instance, paper metrics, paper metadata, and public tags. However, upon trying to interact with the paper by creating annotations, emailing to a colleague, and so on, they will be presented with a log in or sign up option; both will present a single input box for the user's email account. Upon submit, the server will send a key to the user's inbox that expires within 24 hours and can be used on a one-time basis to store a cookie in the user's browser, thus signalling that they are authenticated. Their email account will be hashed and used as their ID.

Users will be able to share annotations privately with other users by emailing the intended recipients directly with either a PDF snippet or all of their annotations. The new user - if they do not have an account - will be able to access these annotations immediately by either clicking a link with

the instance hash on the end which shows that they got it from their email accounts, or by logging in.

5. ACKNOWLEDGEMENTS

This work has been supported by a MetaKnowledge Grant and by MIT.